DOCUMENT RESUME

ED 293 844 TM 011 252

AUTHOR	Kippel, Gary M.; Forehand, Garlie A.
TITLE	SMRT-STEPS: School Mastery of Reading Test System To Enhance Progress of Schools. Fall 1987, Progress Report.
INSTITUTION	New York City Board of Education, Brooklyn. Office of Educational Assessment.
PUB DATE	[87]
NOTE	160p.
PUB TYPE	Reports - Descriptive (141) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Elementary Education; Grade 3; Grade 4; Reading Achievement; Reading Improvement; *Reading Tests; *Testing Programs; Test Reliability; Test Validity
IDENTIFIERS	New York (Brooklyn); *School Mastery of Reading Test

ABSTRACT

The School Mastery of Reading Test (SMRT) program was designed to give administrators and teachers information about reading performance and recommendations for improving the instructional program. The project includes development of a valid and reliable reading test based on the New York City curriculum in order to assess the potential linkage between school level diagnosis and prescription to enhance the progress of schools. The SMRT was administered to 1,979 third and 1,810 fourth graders in the fall of 1986 and the spring of 1987. The students were from nine Comprehensive Assessment Report (CAR) elementary schools in the three Community School Districts. Improvement in the students' scores supports the test's validity. Reliability estimates indicate that the SMRT can be reliably used. A framework for establishing SMRT performance standards is illustrated using the Metropolitan Achievement Test and the Degrees of Reading Power Test with judgments from a panel of New York City educators. Test performances are documented in numerous tables and graphs. Instructions for completing a test review and SMRT-STEPS supplemental guidelines for test review are appended. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Gary M. Kippel

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

SMRT-STEPS

SCHOOL MASTERY OF READING TEST SYSTEM TO ENHANCE PROGRESS OF SCHOOLS

Fall 1987, Progress Report

SMRT-STEPS

SCHOOL MASTERY OF READING TEST
SYSTEM TO ENHANCE PROGRESS OF SCHOOLS

Fall 1987, Progress Report

Report Prepared by:

Gary M. Kippel
Project Director
Office of Educational Assessment
Board of Education of New York City

Garlie A. Forehand
Director of Research Program
Planning and Development
Educational Testing Service
Princeton, New Jersey

New York City Public Schools
Office of Educational Assessment
Richard Guttenberg, Director

It is the policy of the Board of Education not to discriminate on the basis of race, creed, national origin, age, handicapping condition, sexual orientation, or sex, in its educational programs, activities, and employment policies, as required by law. Any person who believes he or she has been discriminated against should contact: Carole Guerra, Local Equal Opportunity Coordinator, Office of Educational Assessment, 110 Livingston Street, Room 743, Brooklyn, New York 11201. Inquiries regarding compliance with appropriate laws may also be directed to: Mercedes A. Nesfield, Director, Office of Equal Opportunity, 110 Livingston Street, Room 601, Brooklyn, New York; or the Director, Office of Civil Rights, U.S. Department of Education, 26 Federal Plaza, Room 33-130, New York, New York 10278.

NEW YORK CITY BOARD OF EDUCATION

Robert F. Wagner, Jr., President

Irene Impellizzeri, Vice President

Gwendolyn C. Baker

Richard I. Beattie

Stephen R. Franse

James F. Regan

Edward L. Sadowsky

Nathan Quinones, Chancellor

Charles I. Schonhaut, Deputy Chancellor

SMRT-STEPS: School Mastery of Reading Test
System To Enhance the Progress of Schools

EXECUTIVE SUMMARY

The primary objective of this project (implemented with the Educational Testing Service of Princeton, New Jersey) is to develop a system to provide school administrators and teachers with information regarding reading performance and recommendations for improving the school instructional program. When the SMRT-STEPS (pronounced: "smart steps") project is complete, it may be considered a school level diagnostic-prescriptive system. The unique and innovative aspects of this project include the development of a valid and reliable test of reading based on the New York City curriculum providing mastery criteria and prescriptive guidelines.

In both fall 1986 and spring 1987, the School Mastery of Reading Test (SMRT) was administered to third and fourth graders in nine Comprehensive Assessment Report (CAR) elementary schools in three Community School Districts. The following provides a summary of findings and accomplishments:

- In both grades three and four, scores from the spring 1987 SMRT administration were consistently higher than scores from the fall 1986 test administration. In addition, grade four test scores were generally higher than those for grade three. These findings suggest the validity of SMRT. Also, differences between cross-sectional and longitudinal data were observed and reported.

- Third and fourth grade students obtained highest percentage of items correct on the word attack subtest and lowest on the reasoning comprehension subtest. This is consistent with curriculum and instruction emphasis

- In both grades three and four, test score distributions especially in spring were negatively skewed indicating a "piling up of scores" at the high end of the score distribution. This is the type of test score distribution which would be expected from a mastery test related to curriculum and administered at the end of the academic year

- Correlational evidence supports the validity of the SMRT subtests

- Grades three and four reliability estimates, resulting from both fall and spring test administrations, provide support for the contention that SMRT can be used reliably

- The validity of calibrating SMRT items onto the National Assessment of Educational Progress (NAEP) scale has been demonstrated. Consequently, SMRT results can be interpreted with respect to NAEP national norms and performance standards. Furthermore, SMRT items can be replaced with comparable NAEP items

- A framework for establishing SMRT performance standards is illustrated using the Metropolitan Achievement Test (MAT), Degrees of Reading Power (DRP) test data and expert judgments from a professional panel of New York City educators

TABLE OF CONTENTS

CHAPTER	TITLE	PAGE
I.	Brief Description of the School Mastery of Reading Test System to Enhance the Progress of Schools (SMRT-STEPS) Project	1
II.	Unique and Innovative Aspects of the School Mastery of Reading Test System to Enhance the Progress of Schools (SMRT-STEPS) Project	3
III.	The School Mastery of Reading Test (SMRT)	4
IV.	The Relationship Between the School Mastery of Reading Test (SMRT) and New York City Curriculum	10
V.	Test Administration	12
VI.	Spring 1987 Results and Longitudinal Comparisons	16
VII.	Reliability of the School Mastery of Reading Test (SMRT) for Grades Three and Four	62
VIII.	Development of Subtests for the School Mastery of Reading Test (SMRT)	65
IX.	The School Mastery of Reading Test (SMRT) and National Assessment of Educational Progress (NAEP) Norms and Performance Standards	72
X.	Establishing Minimum Standards for the School Mastery of Reading Test (SMRT)	85
XI.	Relationship Between SMRT-STEPS and CIMS-CA Project	106
XII.	Review of Other Standardized Reading Tests	109
XIII.	Summary of Findings and Accomplishments	115
Bibliography		118
Appendices		121

LIST OF TABLES

TABLE	TITLE	PAGE
1.	School Mastery of Reading Test Blueprint and Time Guidelines	6
2.	Description of School Mastery of Reading Test Subtests	8
3.	May 1987 Participating Schools and Quantities of Test Materials	14
4.	Fall 1986 and Spring 1987 Delivery and Retrieval Schedule	15
5.	Fall 1986 and Spring 1987 School Mastery of Reading Test Means for Grade Three	20
6.	Fall 1986 and Spring 1987 School Mastery of Reading Test Medians for Grade Three	21
7.	Fall 1986 and Spring 1987 School Mastery of Reading Test Subtest Means for Grade Three	22
8.	Fall 1986 and Spring 1987 School Mastery of Reading Test Subtest Medians for Grade Three	23
9.	Longitudinal Fall 1986 and Spring 1987 School Mastery of Reading Test Means for Grade Three	24
10.	Longitudinal Fall 1986 and Spring 1987 School Mastery of Reading Test Medians for Grade Three	25
11.	Fall 1986 and Spring 1987 School Mastery of Reading Test Means for Grade Four	36
12.	Fall 1986 and Spring 1987 School Mastery of Reading Test Medians for Grade Four	37
13.	Fall 1986 and Spring 1987 School Mastery of Reading Test Subtest Means for Grade Four	38
14.	Fall 1986 and Spring 1987 School Mastery of Reading Test Subtest Medians for Grade Four	39
15.	Longitudinal Fall 1986 and Spring 1987 School Mastery of Reading Test Means for Grade Four	40
16.	Longitudinal Fall 1986 and Spring 1987 School Mastery of Reading Test Medians for Grade Four	41

LIST OF TABLES (continued)

TABLE	TITLE	PAGE
17.	Spring 1987 School Mastery of Reading Test Means for Grades Three and Four	51
18.	Spring 1987 School Mastery of Reading Test Medians for Grades Three and Four	52
19.	Spring 1987 School Mastery of Reading Test Subtest Means for Grades Three and Four	53
20.	Spring 1987 School Mastery of Reading Test Subtest Medians for Grades Three and Four	54
21.	Reliability of the School Mastery of Reading Test Fall 1986 and Spring 1987 for Grade Three	63
22.	Reliability of the School Mastery of Reading Test Fall 1986 and Spring 1987 for Grade Four	64
23.	Correlations Between the School Mastery of Reading Test (SMRT) Subtests	68
24.	Correlations Between Modified Part I and Modified Part II Subtests of the School Mastery of Reading Test (SMRT)	69
25.	Correlations Between Modified Subtests Within Part I of the School Mastery of Reading Test (SMRT)	70
26.	Correlations Between Modified Subtests Within Part II of the School Mastery of Reading Test (SMRT)	71
27.	Percentage of Correct Items for the National Assessment of Educational Progress (NAEP) and the School Mastery of Reading Test (SMRT)	79
28.	National Assessment of Educational Progress (NAEP) Levels of Proficiency	80
29.	Summary of Actual and Expected School Mastery of Reading Test (SMRT) Performance for Below Minimal Competence, Minimally Competent and Competent Readers	90
30.	Citywide Numbers of Grade Four Students in Each of Three Categories of Competence	91
31.	Number and Percentage of Professional Panel Judgments of the Proportion of Students Expected to Respond Correctly on the (97 Item) SMRT	92

LIST OF TABLES (continued)

TABLE	TITLE	PAGE
32.	Number and Percentage of Professional Panel Judgments of the Proportion of Students Expected to Respond Correctly on the SMRT Word Attack (18 Item) Subtest	93
33.	Number and Percentage of Professional Panel Judgments of the Proportion of Students Expected to Respond Correctly on the SMRT Word Meaning (21 Item) Subtest	94
34.	Number and Percentage of Professional Panel Judgments of the Proportion of Students Expected to Respond Correctly on the SMRT Literal Comprehension (31 Item) Subtest)	95
35.	Number and Percentage of Professional Panel Judgments of the Proportion of Students Expected to Respond Correctly on the SMRT Reasoning Comprehension (27 Item) Subtest	96
36.	School Mastery of Reading Test (SMRT) Panel Expectations For Low, Marginal and High Scoring Groups	97
37.	School Mastery of Reading Test (SMRT) Performance for Three Degrees of Reading Power (DRP) Groups	98
38.	School Mastery of Reading Test (SMRT) Performance for Three Metropolitan Achievement Test (MAT) Groups	99
39.	Number and Percent of Correct Responses to Each School Mastery of Reading Test (SMRT) Item for Three Groups of Students Defined by their Metropolitan Achievement Test (MAT) Scores	100
40.	The Relationship Between SMRT-STEPS and CIMS-CA	107
41.	Overview of Frequently Used Standardized Reading Tests	114

LIST OF FIGURES

FIGURE	TITLE	PAGE
1.	School Mastery of Reading Test Proportion of Subtest Items (Pie Chart)	9
2.	Longitudinal Fall 1986 and Spring 1987 Grade 3 Citywide Results (Histogram)	26
3.	Longitudinal Fall 1986 and Spring 1987 Grade 3 District 17 Results (Histogram)	27
4.	Longitudinal Fall 1986 and Spring 1987 Grade 3 District 19 Results (Histogram)	28
5.	Longitudinal Fall 1986 and Spring 1987 Grade 3 District 21 Results (Histogram)	29
6.	Longitudinal Fall 1986 and Spring 1987 Grade 3 Citywide Total Test Results (Graph)	30
7.	Longitudinal Fall 1986 and Spring 1987 Grade 3 Citywide Subtest Results (Graph)	31
8.	Longitudinal Fall 1986 and Spring 1987 Grade 3 Total Test Results For Three Community School Districts (Graph)	32
9.	Longitudinal Fall 1986 and Spring 1987 Grade 4 Citywide Results (Histogram)	42
10.	Longitudinal Fall 1986 and Spring 1987 Grade 4 District 17 Results (Histogram)	43
11.	Longitudinal Fall 1986 and Spring 1987 Grade 4 District 19 Results (Histogram)	44
12.	Longitudinal Fall 1986 and Spring 1987 Grade 4 District 21 Results (Histogram)	45
13.	Longitudinal Fall 1986 and Spring 1987 Grade 4 Citywide Total Test Results (Graph)	46
14.	Longitudinal Fall 1986 and Spring 1987 Grade 4 Citywide Subtest Results (Graph)	47
15.	Longitudinal Fall 1986 and Spring 1987 Grade 4 Total Test Results For Three Community School Districts (Graph)	48
16.	Spring 1987 Grades 3 and 4 Citywide Results (Histogram)	55

LIST OF FIGURES (continued)

FIGURE	TITLE	PAGE
17.	Spring 1987 Grades 3 and 4 District 17 Results (Histogram)	56
18.	Spring 1987 Grades 3 and 4 District 19 Results (Histogram)	57
19.	Spring 1987 Grades 3 and 4 District 21 Results (Histogram)	58
20.	Spring 1987 Grades 3 and 4 Citywide Total Test Results (Graph)	59
21.	Spring 1987 Grades 3 and 4 Citywide Subtest Results (Graph)	60
22.	Spring 1987 Grades 3 and 4 Total Test Results For Three Community School Districts (Graph)	61
23.	Plot of Test Characteristic Curves Across Theta for 91 SMRT Items and 16 NAEP Items	81
24.	Relative Stability of SMRT Item Difficulty Estimates (b-values) From NAEP Precalibrations to New SMRT Estimates	82
25.	Relative Stability of SMRT Item Discrimination Estimates (a-values) From NAEP Precalibrations to New SMRT Estimates	83
26.	Schematic of Horizontal Equating Design for SMRT in terms of NAEP	84
27.	Percent of Total School Mastery of Reading Test Items Correct for Low, Marginal and High Score Groups	103
28.	Percent of Word Attack Subtest Items Correct for Low, Marginal and High Score Groups	104
29.	Percent of Word Meaning Subtest Items Correct for Low, Marginal and High Score Groups	104
30.	Percent of Literal Comprehension Subtest Items Correct for Low, Marginal and High Score Groups	105
31.	Percent of Reasoning Comprehension Subtest Items Correct for Low, Marginal and High Score Groups	105

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
A	Items Obtained From The National Assessment Of Educational Progress (NAEP)	121
B	Instructions For Completing A Test Review	122
C	SMRT-STEPS Supplemental Guidelines For Review Of Tests	125

I. BRIEF DESCRIPTION OF THE SCHOOL MASTERY OF READING TEST SYSTEM TO ENHANCE THE PROGRESS OF SCHOOLS (SMRT-STEPS) PROJECT[1]

The primary objective of this project is to develop a system to provide school administrators and teachers with reading performance scores and information useful for improving the school instructional program. Furthermore, it is our intention to assess the potential linkage between school level diagnosis and prescription in order to enhance the progress of schools. This system is expected to be a particulary useful adjunct to the New York State Comprehensive Assessment Report (CAR) by diagnosing school needs for particular improvement plans developed by the New York City Board of Education.

Consequently, when SMRT-STEPS (pronounced "smart steps") is completely validated it may be considered a school level diagnostic-prescriptive system. In effect, weaknesses requiring remediation will be identified. Subsequently, results from testing may "elicit" or assist in the selection of school improvement plans or corrective actions designed to improve the effectiveness of the instructional program.

To expedite communication, the acronym "SMRT-STEPS" will be used to refer to the entire School Mastery of Reading Test System to Enhance Progress of Schools. The acronym "SMRT" will be used to refer primarily to the assessment component, School Mastery of Reading Test.

To enhance its relevance and usefulness for improving instruction, SMRT is being developed as an objective test of mastery of reading, rather than as a norm-referenced test. As such, SMRT is being designed to indicate the extent to which specific reading skills have been mastered, rather than to differentiate or discriminate between children. Consequently, resulting subtest scores will reflect mastery or competence. This is in contrast to norm-referenced scores such as grade equivalents, normal curve equivalents (NCE's) and percentiles which can be misleading and are susceptible to misinterpretation. It is proposed that the SMRT mastery scores identify separately reported and potentially diagnostic dimensions including word attack, word meaning, literal comprehension, and reasoning comprehension.

---

[1] The assistance of the following SMRT-STEPS Project staff is gratefully acknowledged: K.R. Shivakumar - Education Analyst (SMRT-STEPS Computer Systems Specialist), Charisse Wynn - Associate Word Processor.

The currently available partially validated research version of this test is designed to identify reading subtest areas in which either small instructional groups, intact classes or the entire grade in particular schools are not achieving mastery. The short-term objective is to develop an instructionally useful grade four reading test. The current version has been administered, also, to grade three students. It is anticipated that fall-administered SMRT tests would be most useful to schools for instructional purposes. It would be possible, also, to administer SMRT at various subsequent times throughout the school year to assess progress.

Need for SMRT-STEPS is particularly timely in light of requirements of Part 100 of Commissioner of Education Regulations (New York State Education Department, 1984, 1985). These regulations initiate an innovative Comprehensive Assessment Report (CAR) which summarizes state testing program results, in addition to other school data (e.g., enrollment numbers, graduation results, attendance and dropout rates). Based upon the CAR, 393 New York City Schools (237 elementary, 102 junior high/intermediate and 54 high schools) have been identified by the New York State Education Department as in need of improvement. The primary objective of the SMRT-STEPS Project is to establish a diagnostic-prescriptive system to assist New York City teachers and administrators to improve student reading achievement in such schools. Moreover, SMRT-STEPS will provide information vital for effective planning and policy decisions.

Before implementing the SMRT-STEPS Project, other instructional programs and frequently used standardized reading tests were surveyed. As indicated later in this report (see: "Relationship Between SMRT-STEPS and CIMS-CA Project"), SMRT-STEPS and CIMS-CA differ in nature and scope. Furthermore, among standardized reading tests reviewed and discussed later in this report (see: "Review of Other Standardized Reading Tests"), no existing test was found to be an adequate substitute for a new test based specifically upon New York City curriculum.

II. UNIQUE AND INNOVATIVE ASPECTS OF THE SCHOOL MASTERY OF READING TEST SYSTEM TO ENHANCE THE PROGRESS OF SCHOOLS (SMRT-STEPS) PROJECT

This school improvement system is characterized by the following unique and innovative aspects:

1) It is being developed by a consortium comprised of the New York City Board of Education and the Educational Testing Service of Princeton, New Jersey. In addition to providing a technically sound and useful system, the public schools will not have to pay royalties to a test publisher for the diagnostic part of the system

2) A professional panel of New York City school administrators, teachers, reading experts and curriculum specialists has reviewed SMRT for appropriateness, usefulness and potential bias. This panel will continue to be involved in the program in order to review and establish the relationship between assessment and school improvement materials, plans and programs

3) Common scaling between SMRT and National Assessment of Educational Progress (NAEP) is being established. It is anticipated that SMRT results may be interpreted with respect to NAEP national norms and performance standards. To some extent, also, NAEP might be a cost-effective source of new test items for SMRT

4) The diagnostic component provides an objective test of mastery rather than a norm-referenced test. As such, it is designed to assess reading proficiency and provides a relatively sensitive measure of instruction

5) The diagnostic component is based upon New York City curriculum and provides instructionally useful subscale scores to identify specific reading skills for diagnostic-prescriptive school improvement purposes

6) It is our intention to assess the feasibility of employing advanced computer technology and state-of-the-art psychometric techniques in the development and production of the diagnostic component and, also, in the linkage between assessment and school improvement materials, plans and programs

7) It is our eventual intention to design meaningful and useful reports of test results. Furthermore, the feasibility of relating subtest score profiles to prescriptive choices or menus of school improvement materials, plans and programs will be explored

III. THE SCHOOL MASTERY OF READING TEST (SMRT)

The primary short term objective is to develop SMRT as a standardized measure of reading performance which can be readily administered and scored on a large scale and which accurately reflects multiple skills involved in reading. To enhance its relevance and usefulness for improving instruction, SMRT is being developed as an objective test of mastery of reading, rather than as a norm-referenced test. As such, SMRT is being designed to indicate the extent to which specific reading skills have been mastered, rather than to differentiate or discriminate between children. Consequently, resulting subtest scores will reflect mastery or competence.

As indicated in Table 1A, the current 100 item SMRT consists of four subtests including: word attack (18 items), word meaning (21 items), literal comprehension (31 items), and reasoning comprehension (27 items). When scored, SMRT provides four subtest scores and one total test score. Three additional word recognition items appear at the beginning of the test. These low difficulty items are used to orient students to test directions, format and the separate answer sheet. In addition, they begin students on a positive note in that they are relatively easy items. Descriptions of the different subtests is provided in Table 2. The proportion of items in the four subtests is depicted in Figure 1.

The SMRT booklet is not comprised of clearly defined subtests. Rather, items from the various subtests appear in both parts one and two. Furthermore, the actual tasks required of students change frequently. The specific item numbers for the items in each subtest is presented in Table 1B.

SMRT is divided into two 50 item parts administered with a brief intermission. SMRT is being developed as a "power" test without time limits rather than as a "speed" test. However, time guidelines are provided. As indicated in Table 1C, approximately 33 and 34 minutes are required for administration of parts one and two, respectively, for a total testing time of approximately 67 minutes. In general, most students finish within this time interval.

It is noted that there are a small number of additional SMRT items, including some cloze comprehension items, which were eliminated from the current test in order to limit the amount of time required for test administration. These additional items remain part of the available item bank.

In the current 100 item test, three items are used as "examples" to illustrate directions. These include two word recognition and one word attack item. In effect, students are told the correct answer after they attempt to respond.

Directions are read to students. Incorporated within the remaining 97 items, are 16 items obtained from the National Assessment of Education Progress (NAEP). Of these 16, ten are literal comprehension and the remaining six are reasoning comprehension items. The reasons for embedding NAEP items within SMRT are discussed in the section entitled "The School Mastery of Reading Test (SMRT) and National Assessment of Educational Progress (NAEP) Norms and Performance Standards".

Students respond to test questions on machine scannable general purpose NCS answer sheets (i.e., NCS Trans-Optic EB08-4521:223222). Subsequently, these answer sheets are scanned (see, for discussion of answer key, Kippel and Forehand, 1987, pp. 39-40) on an NCS 7018 Optical Mark Reader with NCS Scanpak "Test Scoring Package" software (see, for discussion of machine scoring procedures, Kippel and Forehand, 1986, pp. 14-15).

Table 1

SCHOOL MASTERY OF READING TEST BLUEPRINT AND TIME GUIDELINES

TABLE 1A: Quantity Of Items

Part or Total	Word Recognition	Word Attack	Word Meaning	Literal Comprehension	Reasoning Comprehension	Total
Part One	3	9	8	15	15	50
Part Two	—	9	13	16	12	50
Total	3	18	21	31	27	100

TABLE 1B: Item Numbers

Part or Total	Word Recognition	Word Attack	Word Meaning	Literal Comprehension	Reasoning Comprehension
Part One (Items 1-50)	1-3	4-12	13-20	21-23, 25-32, 34, 37, 39, 40	24, 33, 35, 36, 38, 41-50
Part Two (Items 51-100)		51-59	60-72	73-80, 83-87 92, 94, 97	81-82, 88-91, 93, 95, 96, 98-100

TABLE 1C: Time Guidelines (in minutes)

Part or Total	Word Recognition	Word Attack	Word Meaning	Literal Comprehension	Reasoning Comprehension	Total
Part One	3	5	4	12	10	34
Part Two	—	5	7	12	9	33
Total	3	10	11	24	19	67

Table 2

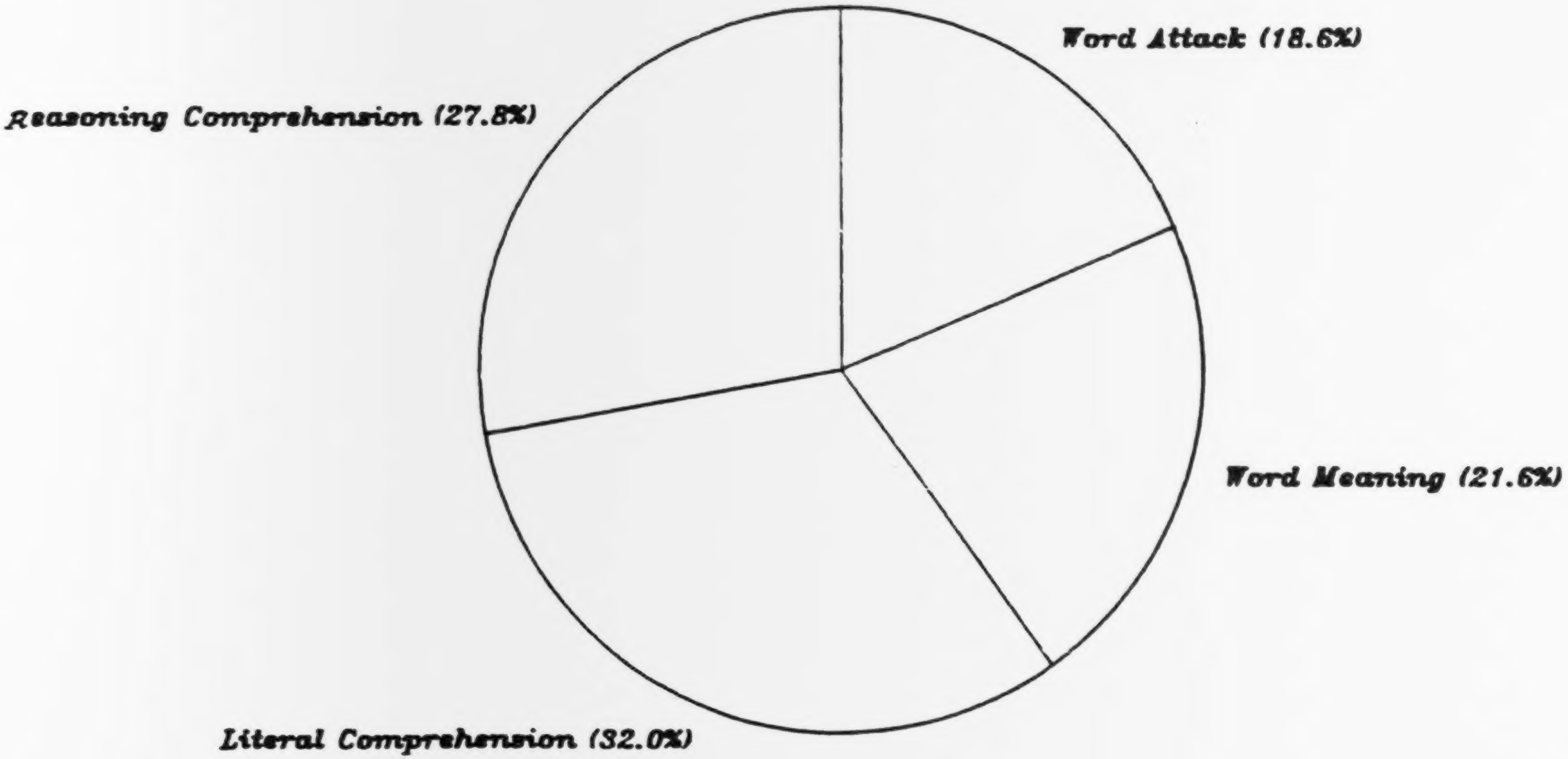
Description of School Mastery of Reading Test Subtests

Category/Subtest	Description
1. Word recognition	The student (1) hears a word and chooses that word from a list of words, (2) reads a word and chooses a matching picture, and (3) looks at a picture and chooses a matching word. The following words are included: of, was, cat, dog, four, from, one, what, some, know, might, flower, night, automobile, piano, birdcage, castle, swords.
2. Word attack	The student (1) hears a word and chooses a word with the same sound from a list of words (ŭ, ā, ō, oi, ow, f, ch, t, gh), and (2) reads a word with a portion underlined and chooses from a list a word with the same sound as the underlined portion (hard c (k), gh, ch, sh, ow, oi (oy), silent b, wr, silent e, soft g).
3. Word meaning	The student (1) matches words to definitions, (2) chooses synonyms and antonyms for words, and (3) chooses words for blank spaces in sentences. The following words are included: ring, cry, chair, night, above, glad, slow, sick, shut, narrow, big, cent, their, children, men, highest, unlike, retell, lost, hide, enjoyed, seen, worked.
4. Literal comprehension	The student reads a sentence, several sentences, or a short story and (1) chooses a sentence that has the same meaning, (2) chooses a picture that best represents the meaning of what was read, and (3) answers factual questions about what was read by choosing from a list of possible answers. The reading material includes: simple sentences, compound subjects and objects, compound and complex sentences.
5. Reasoning comprehension	The student reads a sentence, several sentences, or a short story and answers inferential questions by choosing from a list of pictures or written answers. The reading materials include single paragraphs, a short story, causal and all/some relationships, predicted outcomes, comparisons and sequencing.
6. Comprehension: cloze	The student reads two long stories (six or seven paragraphs each) with seven words missing in each story. For each missing word, the student chooses from a list of five words the word that best completes the meaning of the story.

Figure 1.

School Mastery of Reading Test

Proportion of Subtest Items*



* The three word recognition items have been eliminated.

IV. THE RELATIONSHIP BETWEEN THE SCHOOL MASTERY OF READING TEST (SMRT) AND NEW YORK CITY CURRICULUM

The School Mastery of Reading Test (SMRT) is related specifically to New York City public school curriculum. The relationship between SMRT and New York City reading and language arts curriculum was assessed both by project staff (see, for discussion, Kippel and Forehand, 1986, pp. 52-53) and by curriculum, language arts and reading specialists of the Division of Curriculum and Instruction of the New York City Board of Education. In addition, after administering SMRT to their students, both third and fourth grade teachers provided their opinion that the majority of test questions correspond well to New York City curriculum (see, for discussion, Kippel and Forehand, 1986, pp. 54-56; 1987, pp.36-38). Furthermore, a Professional Panel of New York City teachers and supervisors provided favorable ratings reflecting their opinions of the usefulness of SMRT (see, for discussion, Kippel and Forehand, 1986, pp. 48-49).

Curriculum objectives may be delineated in curriculum guides, or may be inferred from textbooks, workbooks and other instructional materials. For example, in order to establish the congruence between SMRT and New York City public school fourth grade reading curriculum, the following three Board of Education of New York City (1968, 1969, 1980) publications were used to define curriculum: Minimum Teaching Essentials - Grade 3-5, Sequential Levels of Reading Skills, and the Handbook for Language Arts - Grades 3 and 4. In addition, guidance and assistance were provided by citywide curriculum specialists from the Division of Curriculum and Instruction.

The curriculum validity of a test refers to how effectively test objectives represent curriculum objectives. Instructional validity refers to how effectively curriculum objectives were actually taught (see, for discussion, McClung, 1978). To expedite discussion, it is assumed that curriculum objectives are reflected in classroom instruction. In effect, instructional validity is assumed.

To provide instructionally useful information, test objectives must reflect curriculum objectives. In other words, a close match or alignment between test objectives and curriculum objectives is necessary to ensure that test results can be used to improve instructional effectiveness and reading achievement. The optimal decision is to select the standardized test that best matches the curriculum objectives (Wilson & Hiscox, 1984). This makes good sense and is fair. In addition, it may avoid costly litigation if the test is used either to hold over students or to evaluate teacher performance and the test is found, subsequently, not to adequately reflect curriculum.

It is reasonable to assume that there are both similarities and differences between the curricula taught in different school systems throughout the United States of America. Curricula in general may, for example, reflect universal and relatively invariant human growth processes and common curriculum elements which reflect the "state of art" in particular disciplines. In contrast, unique curriculum aspects may reflect locally meaningful curriculum (e.g., New York City geography, history and demographics). Consequently, if the same standardized achievement test was administered in different school systems, that test might be a more effective measure in some school systems compared with others. In other words, some school systems are more likely than others to find a better match or alignment between the objectives included in any specific commercially developed test and that particular school system's curriculum objectives.

There is research evidence of both similarities and differences between curriculum areas assessed between states. For example, Komoski (1987) has analyzed the mathematics curriculum and state test contents in California, Pennsylvania and Tennessee. It is apparent, that there are substantial differences between states. Current additional research by Komoski at the Educational Products Information Exchange (EPIE) is focused on reading and language arts.

In light of curriculum differences between school systems, one may assume that commercial test publishers are likely to develop standardized tests based upon the general or most common aspects of a curriculum rather than tailoring test objectives specifically to the curriculum objectives of any particular school system. In effect, this strategy focuses on a potential regional or nationwide market. It may not, in fact, be financially feasible for a commercial test publisher to limit its potential market by developing valid and reliable standardized tests for any one school system. Consequently, it seems very unlikely that there will be a perfect match between the test objectives and the curriculum of any given school system.

In order to obtain an optimally useful and instructionally meaningful standardized test, it may be both desirable and feasible for the New York City public school system to consider developing its own standardized tests. In addition to being prudent, such a strategy may be cost-effective. Significant savings may result from not being required to pay licensing and royalty fees to commercial test publishers. As discussed in this report, the SMRT-STEPS Project has demonstrated the feasibility of developing a prototype New York City curriculum-based reading test.

V. TEST ADMINISTRATION

Three schools in each of three Brooklyn Community School Districts (see Table 3) participated in both the fall 1986 and the spring 1987 SMRT-STEPS testing program. Each of these schools previously had participated in the spring 1986 SMRT administration. Furthermore, each school had been identified by the New York State Education Department's Comprehensive Assessment Report (CAR) as in need of improvement. Profiles of each of the nine participating schools were presented in Kippel and Forehand (1987, pp 6-8). All participating schools were selected from within the borough of Brooklyn for logistical, control and test security reasons.

During fall 1986 and spring 1987, SMRT was administered in both grades three and four. Specifically, from May 11 through 22, 1987, SMRT was administered to 1,004 grade three and 889 grade four students. Previously, from October 20 through 30, 1986, a total of 975 third grade and 921 fourth grade students were tested.

Schools were requested to complete sheets for every student who was eligible for the annual citywide reading test, with the exception of those Special Education students for whom some testing variance (e.g., large print, extended time limits, etc.) was required. Limited English Proficient students exempted from the annual citywide reading test also were exempted from SMRT. In order to minimize disruption of instruction, provision was not made for "make-up" testing of absentees.

For the spring 1987 test administration, SMRT-STEPS project staff entered student names and nine-digit identification numbers on each machine-scorable answer sheet before they were mailed to the schools. This was done to minimize clerical work required of school personnel.

Test booklets and administration manuals were delivered and retrieved from all nine participating schools by the same companies that transport citywide test materials. The schedule depicted in Table 4 was followed.

To ensure test security, each test administration manual and test booklet was stamped with a unique identification number (see Table 3). Careful track was kept of the range of numbers on both administration manuals and test booklets delivered to, and retrieved from, every school.

All test materials were delivered in strong cartons carefully sealed with white tape with the following message in red letters: "SECURE TEST MATERIALS - DO NOT OPEN." Cartons were delivered directly to the Principal's office and receipts were signed. The sealed cartons were then placed in secure storage closets, usually in the Principal's office.

On the day of testing, project staff visited each school. They retrieved the sealed cartons from locked closets and distributed test materials to participating classes. A careful accounting was maintained of the quantity and identification numbers of both test administration manuals and test booklets delivered to, and subsequently retrieved from, each class.

Project staff monitored the test administration in each school. All tests were administered by third and fourth grade teachers using the test administration manual prepared for that purpose. Appropriate signs were placed on the door of each class indicating that "Testing" was being conducted. Students read the test questions from their test booklet, then responded on the separate machine-scorable answer sheet provided for that purpose. After the fall 1986 testing, each teacher was asked to complete a one-page survey designed to assess their opinions of the test and testing procedures. Results from this survey were reported in Kippel and Forehand (1987, pp. 36-38).

Table 3

May 1987 Participating Schools and Quantities of Test Materials

Community School District Number	School Number	Number of Tests Sent Including Overage	Range of Test Numbers	Range of Admin. Manual Numbers
CSD #17	191	350	1- 350	1- 20
	289	450	351- 800	21- 40
	398	500	801-1,300	41- 60
(Subtotal)		(1,300)		(60)
CSD #19	213	300	1,301-1,600	61- 80
	290	250	1,601-1,850	81-100
	328	150	1,851-2,000	101-120
(Subtotal)		(700)		(60)
CSD #21	90	200	2,001-2,200	121-140
	212	250	2,201-2,450	141-161
	329	200	2,451-2,650	161-180
(Subtotal)		(650)		(60)
[Total]		[2,650]		[180]

Table 4

Fall 1986 and Spring 1987 Test Delivery and Retrieval Schedule

	Fall 1986	Spring 1987
Cartons were obtained from 110 Livingston Street, Room 714, and delivered the same day to all nine schools.	Thursday October 16, 1986	Thursday May 7, 1987
Cartons were retrieved from all nine schools and delivered to the Scan Center, 49 Flatbush Avenue, Brooklyn, New York.	Wednesday November 5, 1986	Tuesday May 26, 1987

VI. SPRING 1987 RESULTS AND LONGITUDINAL COMPARISONS

In the following sections grade three and grade four data are analyzed separately to identify longitudinal trends (i.e., from fall to spring) in each grade. Subsequently, spring 1987 grade three and grade four data are compared.

Grade Three

Review of Tables 5 through 10 reveals that spring 1987 grade three subtest and total test means and medians are consistently higher than those for fall 1987 in all three participating Community School Districts.

Specifically, Tables 5 and 6 present total test means and medians, respectively, for each of the nine participating schools. Tables 7 and 8 present means and medians, respectively, for each of the four SMRT subtests and the total test, for all grade three students tested. Tables 9 and 10 present means and medians, respectively, for longitudinally matched data resulting from the fall 1986 and spring 1987 test administrations.

For all three school districts combined ("grade three citywide"), the grade three total test mean score for fall 1986 is 64.29 with a standard deviation of 18.58 (see Table 5). The corresponding median is 65.00 with a semi-interquartile range of 16.00 (see Table 6). The grade three spring mean and standard deviation are 74.66 and 15.89, respectively. The corresponding median is 78.00 with a semi-interquartile range of 10.88. From these tables it is evident that, in each of the nine participating schools, grade three mean and median SMRT total scores resulting from the spring test administration are higher than the corresponding scores from the fall test administration.

Furthermore, the statistics resulting from the spring test administration, in particular, reflect a negatively skewed distribution. In other words, when administered in the spring, this is a relatively easy test with a "piling up of scores" at the high end of the score distribution. This distribution was expected for a curriculum-based test, such as SMRT, administered in the spring, near the end of the school year. It is likely that the relatively high overall scores reflect mastery, at least to some extent, of third grade curriculum taught during the school year.

Tables 7 and 8, respectively, present the raw score means and medians for each SMRT subtest and the total test. For the fall grade three test administration, the mean number of items answered correctly for the word attack, word meaning, literal comprehension and reasoning comprehension subtests represent 69.78, 62.76, 65.23 and 57.33 percent, respectively, of the items on each subtest. The mean number of the 97 items (i.e., the three word recognition items were not included in these analyses) answered correctly for fall grade three represents 63.32 percent. The corresponding subtest values for the spring grade three test administration are 77.17, 73.95, 76.19 and 69.11. The mean number of 97 items answered correctly in spring is 73.92 percent. For each of the four SMRT subtests and the total test, it is apparent that results from the spring test administration reflect higher achievement than those obtained from the fall test administration.

In order to assess longitudinal trends, the 975 fall 1986 and the 1,004 spring 1987 grade three student answer sheets were matched to obtain 807 pairs of scores from students who attended both fall and spring test administrations. In other words, these data are longitudinal in the sense that each of the 807 students contributed both fall and spring scores. To accomplish the computerized match of fall and spring test scores, unique nine-digit student identification numbers were used. Subsequently, the accuracy of all matched pairs was verified by visual examination of students' names and dates of birth. Grade three means and medians based upon longitudinally matched individual student scores are presented in Tables 9 and 10, respectively.

On the whole, these grade three longitudinal results reported in Tables 9 and 10 closely match the results reported in Tables 5 and 6 which are based upon cross-sectional data. Cross-sectional analyses, in contrast to longitudinal analyses, are not based upon matched pairs of scores (i.e., from the fall and spring administrations) obtained from the same students.

The longitudinally based total test mean for fall 1986 is 65.27 with a standard deviation of 18.39 (see Table 9), as compared with 64.29 and 18.58, respectively, for cross-sectional data (see Table 5). Similarly, the corresponding longitudinally based median is 67.00 with a semi-interquartile range of 15.00 (see Table 10), as compared with 65.00 and 16.00, respectively, for cross-sectional data (see Table 6).

Furthermore, review of the longitudinal results reported in Tables 9 and 10 reveals that, in each of the nine participating schools, grade three mean and median SMRT total scores resulting from the spring administration were higher than the corresponding scores for the fall administration. Again, these trends are consistent with those reported for cross-sectional data (see Tables 5 and 6).

From a research standpoint, longitudinal or "paired comparison" results are more desirable because they provide a more accurate and reliable estimate of improvement in reading performance. The results presented in Tables 9 and 10 (and, subsequently, in Figures 2 through 8) are such longitudinal results, because each student provided pairs of test scores resulting from the fall and spring test administrations.

A close comparison of Tables 5 and 9 reveals certain similarities. First, the rank order of schools within districts 17 and 21 remained constant for both fall and spring test administrations. Second, the rank order of schools was identical for both longitudinal and cross-sectional data. In other words, when looking either at cross-sectional (i.e., Table 5) or longitudinal data (i.e., Table 9), the school that achieved the highest scores within these districts in the fall, also achieved the highest scores in the spring. Third, for each school tested, the increase in mean score from fall to spring administrations was similar for both cross-sectional and longitudinal data.

District 19, however, presented a different picture. Taking the three points in order: First, unlike districts 17 and 21, the rank order of the three schools within district 19, varied from fall to spring. Second, the rank order of schools differed from longitudinal to cross-sectional data. Third, the increase in mean score from fall to spring administrations was greater for longitudinal data than for cross-sectional data. As a result, when comparing the three districts, and considering longitudinal data only, district 19 showed the greatest improvement in mean score from fall to spring (see Table 9). However, when considering the cross-sectional data presented in Table 5 and comparing the three districts, district 17 shows the greatest increase in performance.

Spring cross-sectional results (see, especially, Tables 5 and 6) include scores obtained from students who did not take the test in fall. These include, for example, absentees and students who enrolled after the fall test administration. Such students, especially in P.S. 213 and P.S. 290 (i.e., in school district 19) achieved lower reading scores. Their scores, when combined with those of other students, depress or lower the spring 1987 cross-sectional mean. The results obtained from longitudinal data do not include this group of students and may thus be more representative of the true increase in performance over time in these schools.

Longitudinally matched subtest and total test grade three results for fall and spring are depicted for all three districts combined ("grade three citywide") in Figure 2, and for each of the three participating Community School Districts in Figures 3, 4 and 5. Subsequently, Figures 6 through 8 depict the increase of longitudinally matched grade three spring scores over those for fall. The discrete points shown in these figures have been connected by lines to emphasize the increase. These lines represent an interpolation between the fall and spring scores rather than actual test scores. Furthermore, it is not assumed that increase in achievement is linear.

The difference between fall and spring results for the total test and subtests, respectively, is depicted for all three districts combined ("grade three citywide") in Figures 6 and 7. Specifically, the four subtest scores increased from fall to spring. Further, the students seem to have the greatest difficulty with reasoning comprehension, and perform best on the word attack subtest. From these figures, it is evident that the grade three spring 1987 subtest and total test scores are consistently higher than those for fall 1986 in all three participating Community School Districts.

The difference between fall and spring total test results for each of the three participating Community School Districts is depicted in Figure 8. A comparison of the three districts reveals that all three showed improvement from fall to spring, with district 19 showing the greatest improvement. District 17 was consistently the district with the lowest scores of these three school districts.

TABLE 5

FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST
MEANS FOR GRADE THREE

Community School District (CSD) and Public School (PS)	Fall 1986			Spring 1987		
	Number Of Students	Mean	Standard Deviation	Number Of Students	Mean	Standard Deviation
CSD 17	499	61.84	18.59	489	73.83	15.98
PS 191	128	63.51	17.83	128	75.61	14.67
PS 289	171	67.13	17.49	173	76.18	15.06
PS 398	200	56.25	18.53	188	70.45	17.11
CSD 19	244	65.09	19.40	280	73.84	16.77
PS 213	94	64.72	21.51	116	74.98	15.99
PS 290	96	66.29	19.77	115	71.97	18.57
PS 328	54	63.61	14.43	49	75.51	13.71
CSD 21	232	68.70	16.79	235	77.35	14.31
PS 90	68	70.43	16.27	65	79.85	13.06
PS 212	93	69.52	16.78	95	76.69	15.02
PS 329	71	65.97	17.18	75	76.03	14.34
TOTAL	975	64.29	18.58	1004	74.66	15.89

TABLE 6

FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST MEDIANS FOR GRADE THREE

Community School District (CSD) and Public School (PS)	Fall 1986			Spring 1987		
	Number Of Students	Median	Semi Inter-Quartile Range	Number Of Students	Median	Semi Inter-Quartile Range
CSD 17	499	62.00	16.00	489	77.00	11.00
PS 191	128	63.50	14.38	128	78.00	8.88
PS 289	171	67.00	14.50	173	78.00	9.25
PS 398	200	53.00	14.00	188	72.00	13.50
CSD 19	244	68.50	15.50	280	79.00	11.88
PS 213	94	72.00	19.38	116	79.00	9.88
PS 290	96	71.50	17.75	115	78.00	13.50
PS 328	54	65.00	9.63	49	77.00	7.75
CSD 21	232	70.00	12.50	235	80.00	9.00
PS 90	68	72.50	13.00	65	82.00	7.50
PS 212	93	69.00	12.00	95	79.00	8.50
PS 329	71	69.00	14.50	75	80.00	10.00
TOTAL	975	65.00	16.00	1004	78.00	10.88

TABLE 7

FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST SUBTEST MEANS FOR GRADE THREE

SMRT Subtests	Number of Items*	Fall 1986		Spring 1987	
		Mean	Standard Deviation	Mean	Standard Deviation
Word Attack	18	12.56	3.71	13.89	3.26
Word Meaning	21	13.18	4.84	15.53	4.12
Literal Comprehension	31	20.22	6.54	23.62	5.63
Reasoning Comprehension	27	15.48	5.23	18.66	4.74
Total	97	61.42	18.46	71.70	15.84

* The three word recognition items have been eliminated from these analyses.

TABLE 8

FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST SUBTEST MEDIANS FOR GRADE THREE

SMRT Subtests	Number of Items*	Fall 1986		Spring 1987	
		Median	Semi Inter-Quartile Range	Median	Semi Inter-Quartile Range
Word Attack	18	13.00	3.00	15.00	2.50
Word Meaning	21	14.00	4.00	16.00	3.00
Literal Comprehension	31	21.00	5.50	25.00	4.00
Reasoning Comprehension	27	16.00	4.50	19.00	3.00
Total	97	63.00	16.00	75.00	10.50

* The three word recognition items have been eliminated from these analyses.

TABLE 9

LONGITUDINAL FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST MEANS FOR GRADE THREE

Community School District (CSD) and Public School (PS)	Number Of Students	Fall 1986		Spring 1987	
		Mean	Standard Deviation	Mean	Standard Deviation
CSD 17	409	63.58	18.54	74.62	15.73
PS 191	105	65.83	17.44	76.55	13.62
PS 289	141	68.16	17.55	77.21	14.78
PS 398	163	58.17	18.79	71.14	17.16
CSD 19	199	65.08	19.44	78.38	14.33
PS 213	77	64.71	21.79	79.10	14.70
PS 290	76	65.53	19.99	79.51	14.49
PS 328	46	64.94	13.91	75.28	13.28
CSD 21	199	68.94	16.48	78.45	13.86
PS 90	60	70.95	15.89	80.57	12.15
PS 212	80	69.38	16.36	77.98	15.23
PS 329	59	66.29	17.14	76.93	13.52
TOTAL	807	65.27	18.39	76.49	15.05

TABLE 10

LONGITUDINAL FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST MEDIANS FOR GRADE THREE

Community School District (CSD) and Public School (PS)	Number Of Students	Fall 1986		Spring 1987	
		Median	Semi Inter-Quartile Range	Median	Semi Inter-Quartile Range
CSD 17	409	64.00	16.00	77.00	11.00
PS 191	105	66.00	13.75	78.00	8.00
PS 289	141	70.00	15.00	79.00	8.50
PS 398	163	56.00	15.00	72.00	13.50
CSD 19	199	68.00	15.50	82.00	9.00
PS 213	77	72.00	18.50	83.00	8.00
PS 290	76	69.50	18.88	85.00	9.38
PS 328	46	66.50	9.75	77.00	6.38
CSD 21	199	70.00	12.00	81.00	8.50
PS 90	60	73.00	11.75	82.00	7.25
PS 212	80	68.50	12.00	80.00	9.00
PS 329	59	69.00	13.00	81.00	9.00
TOTAL	807	67.00	15.00	80.00	10.00

Figure 2

Longitudinal Fall 1986 and Spring 1987 Grade 3 Citywide Results

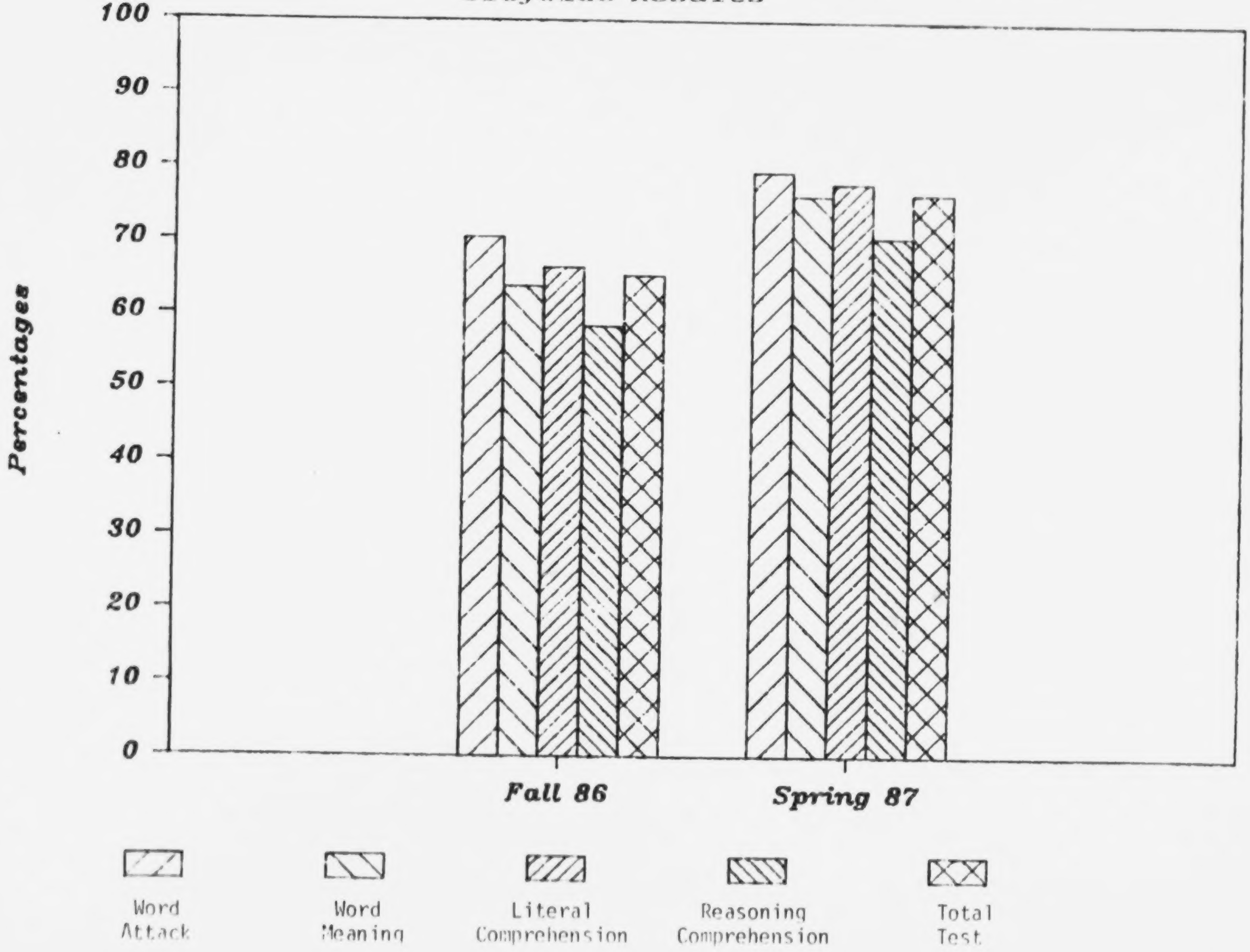


Figure 3

Longitudinal Fall 1986 and Spring 1987 Grade 3 District 17 Results

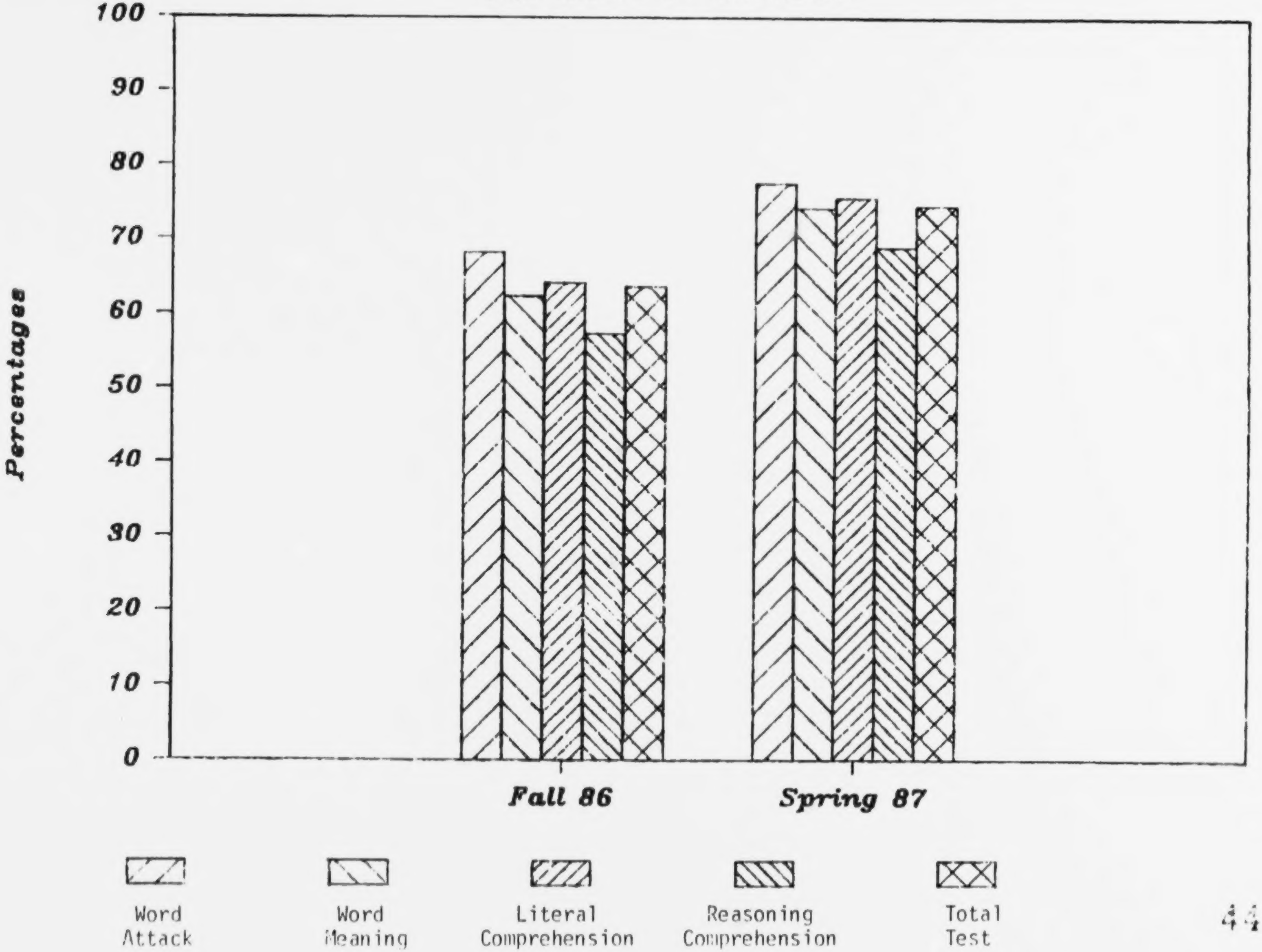


Figure 4

Longitudinal Fall 1986 and Spring 1987 Grade 3 District 19 Results

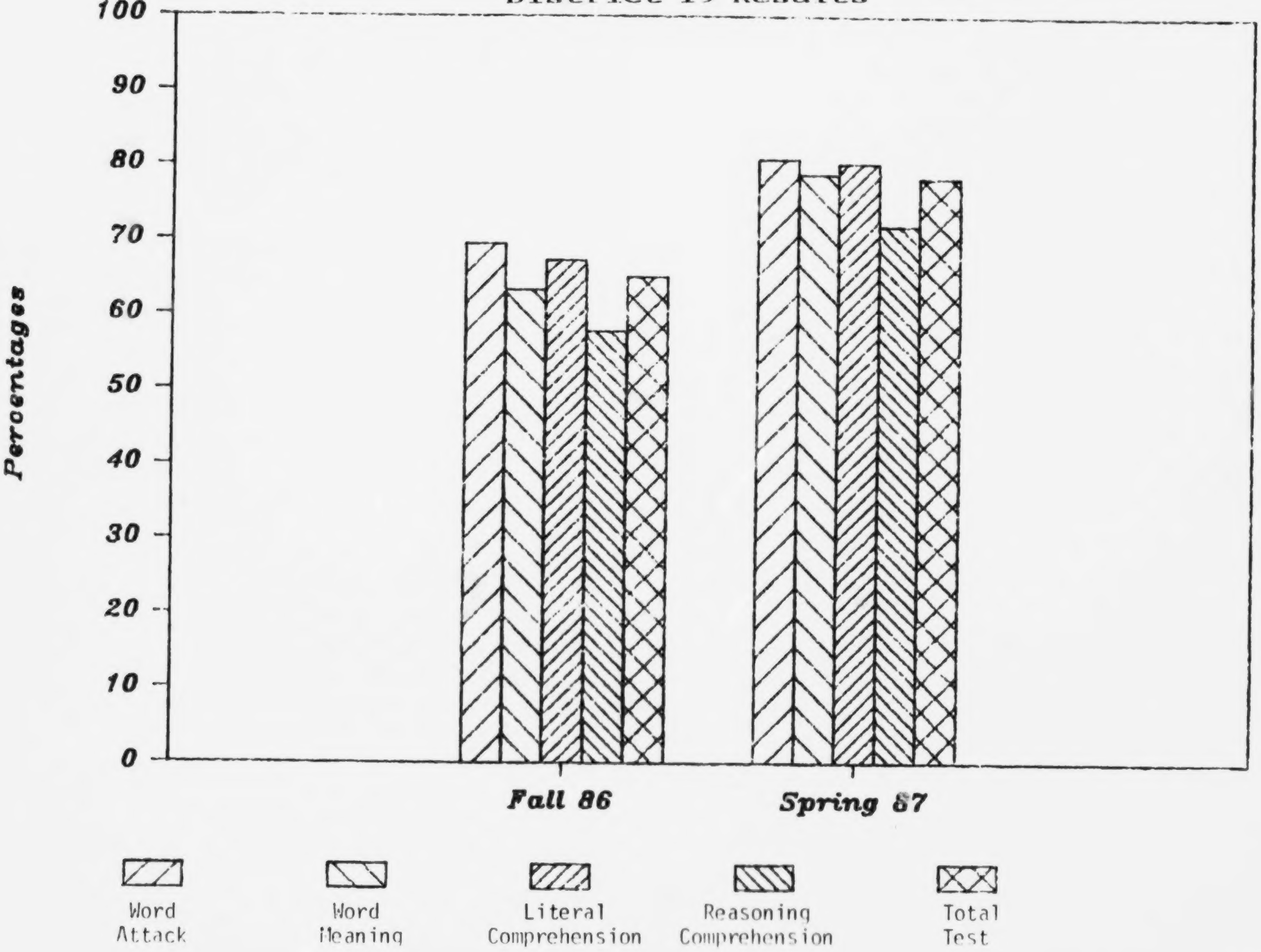


Figure 5

Longitudinal Fall 1986 and Spring 1987 Grade 3
District 21 Results

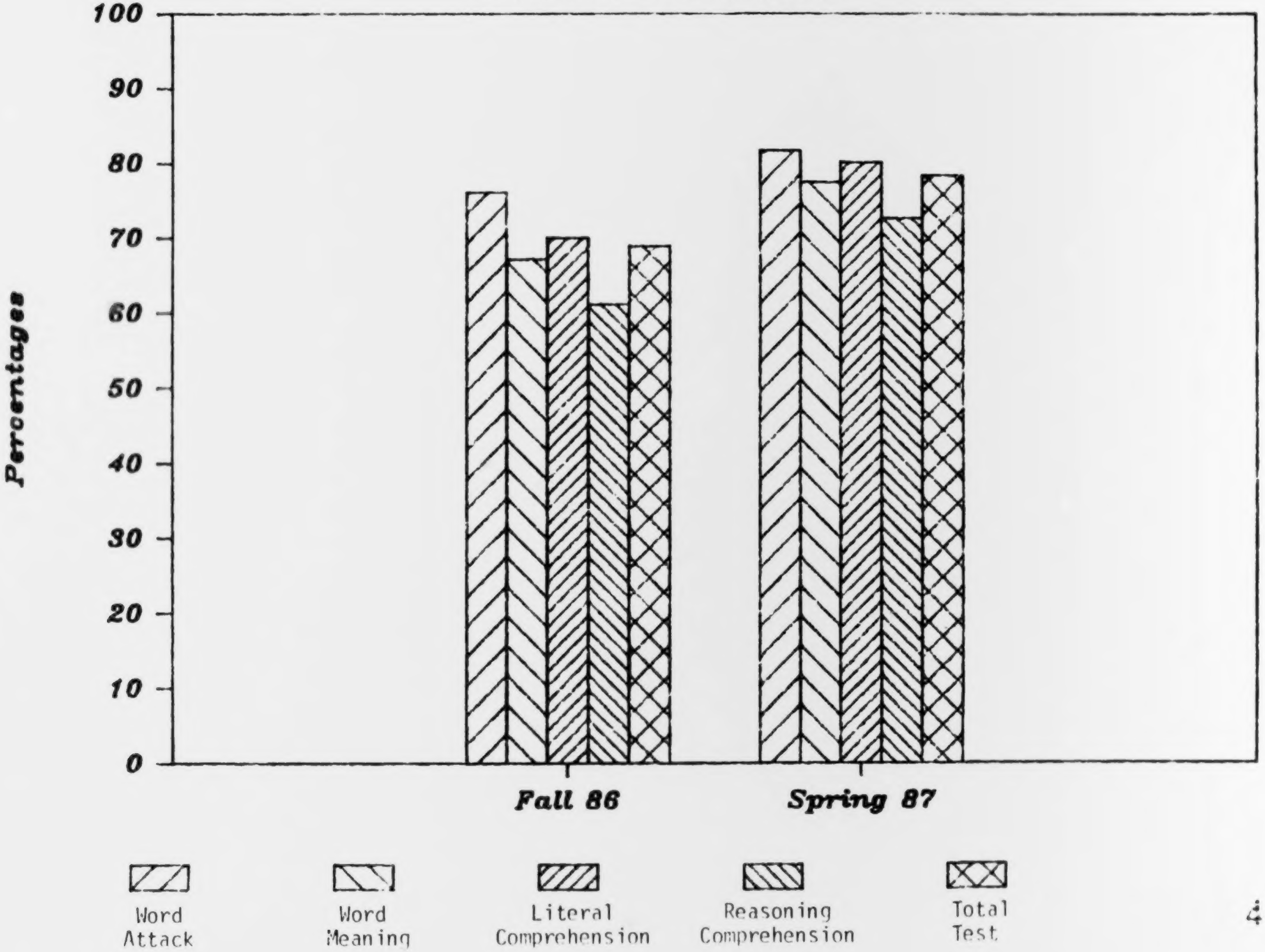


Figure 6

Longitudinal Fall 1986 and Spring 1987 Grade 3 Citywide Total Test Results

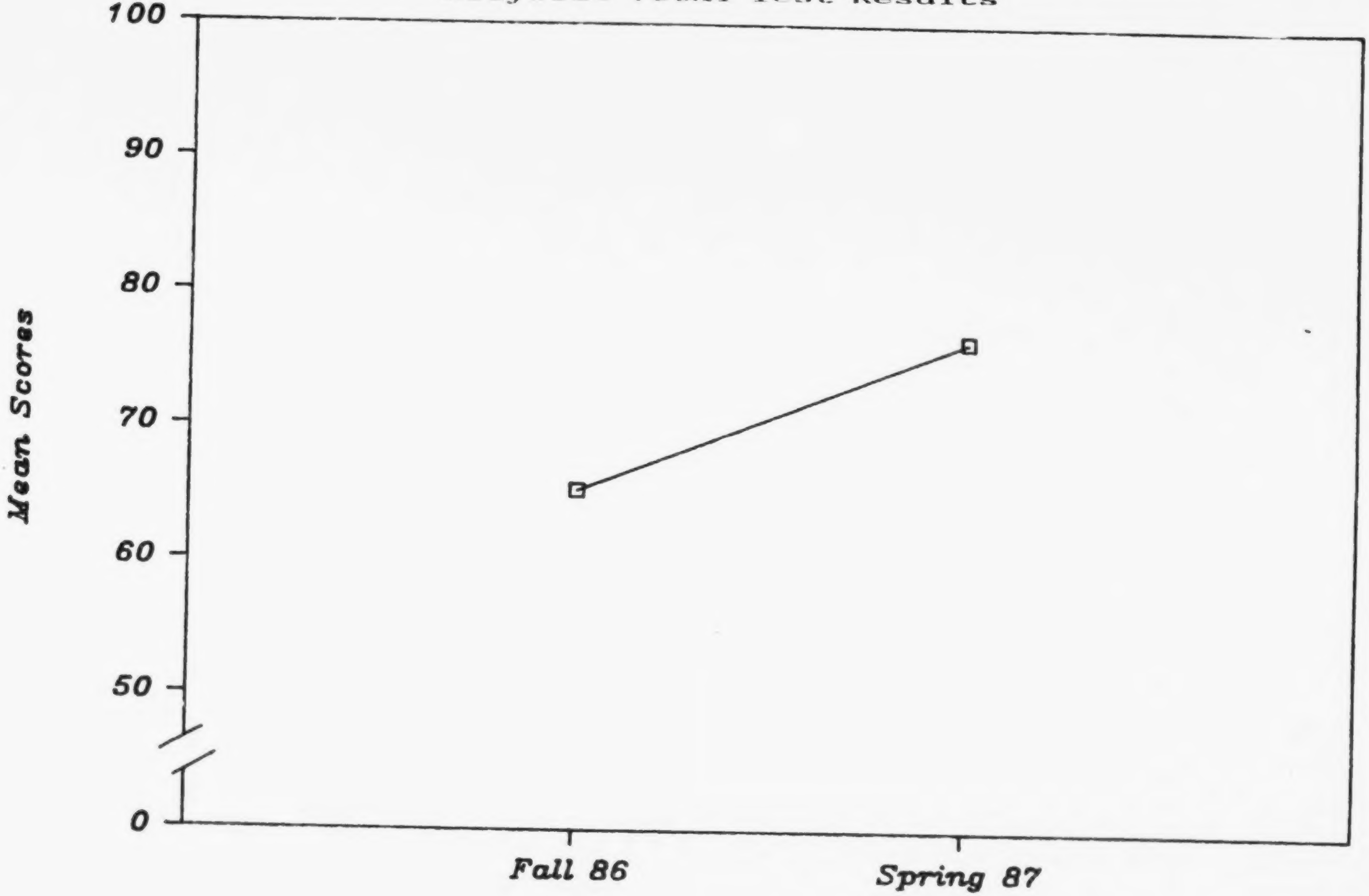


Figure 7

Longitudinal Fall 1986 and Spring 1987 Grade 3 Citywide Subtest Results

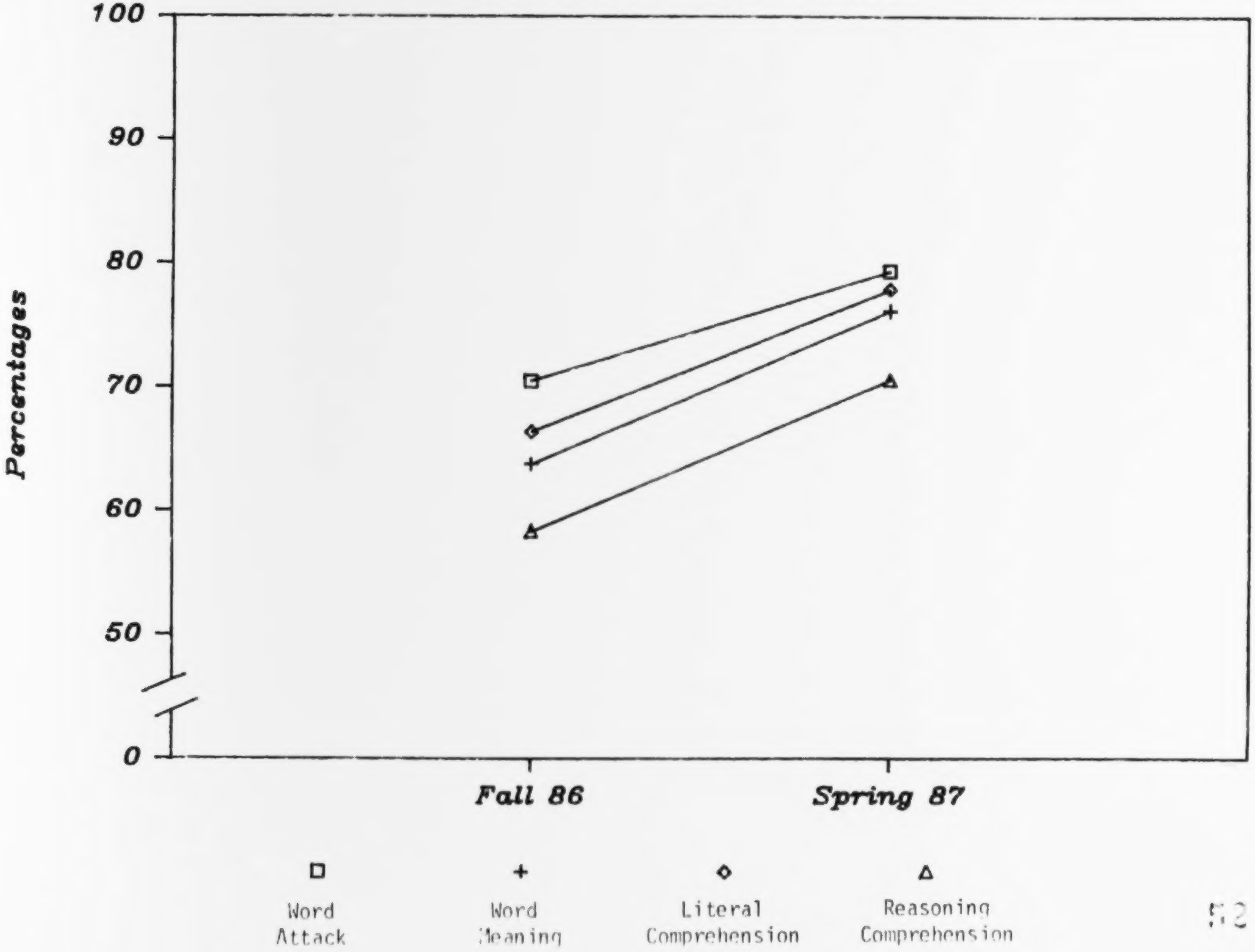
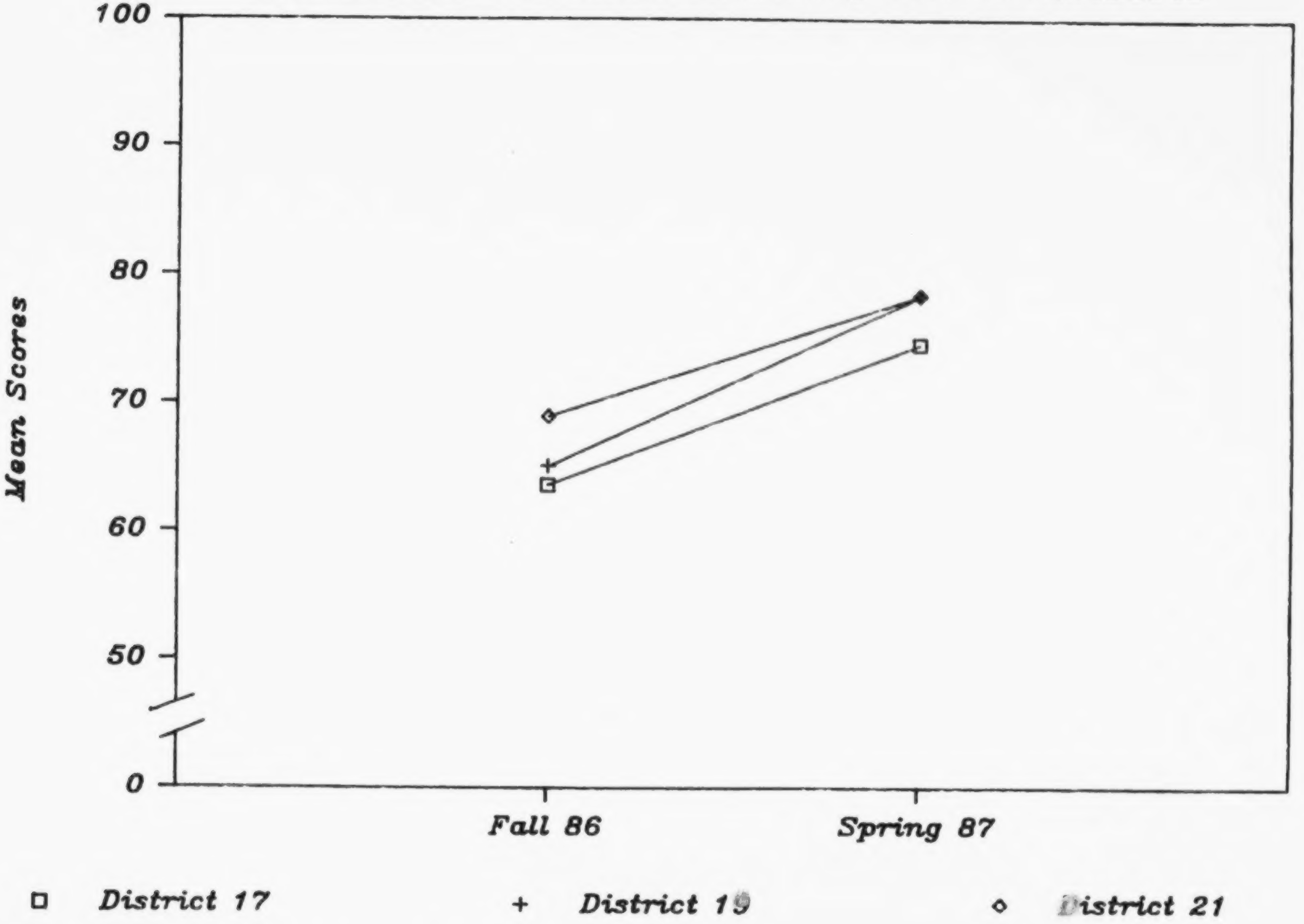


Figure 8

Longitudinal Fall 1986 and Spring 1987 Grade 3 Total Test Results For Three Community School Districts



Grade Four

Review of Tables 11 through 16 reveals that spring 1987 grade four subtest and total test means and medians are consistently higher than those for fall 1986 in all participating Community School Districts.

Specifically, Tables 11 and 12 present total test means and medians, respectively, for each of the nine participating schools. Tables 13 and 14 present means and medians, respectively, for each of the four SMRT subtests and the total test, for all grade four students tested. Tables 15 and 16 present means and medians, respectively, for longitudinally matched data resulting from the fall 1986 and spring 1987 test administrations.

For all three school districts combined ("grade four citywide"), the grade four test mean score for fall 1986 is 75.46 with a standard deviation of 15.22 (see Table 11). The corresponding median is 78.00 with a semi-interquartile range of 9.50 (see Table 12). The grade four spring 1987 mean and standard deviation are 81.20 and 13.91, respectively. The corresponding median is 85.00 with a semi-interquartile range of 8.00. From these tables it is evident that, in each of the nine participating schools, grade four mean and median SMRT total scores resulting from the spring test administration are higher than corresponding scores from the fall test administration.

Furthermore, the statistics resulting from the spring test administration, in particular, reflect a negatively skewed distribution. In other words, when administered in the spring, this is a relatively easy test with a "piling up of scores" at the high end of the score distribution. This distribution was expected for a curriculum-based test, such as SMRT, administered in the spring, near the end of the school year. It is likely that the relatively high overall scores reflect mastery, at least to some extent, of fourth grade curriculum taught during the school year.

Tables 13 and 14, respectively, present the raw score means and medians for each SMRT subtest and the total test. For the fall grade four test administration, the mean number of items answered correctly for the word attack, word meaning, literal comprehension and reasoning comprehension subtests represent 78.83, 75.76, 76.48 and 69.26 percent, respectively, of the items on each subtest. The mean number of the 97 items (i.e., the three word recognition items were not included in these analyses) answered correctly for fall grade four represents 74.75 percent. The corresponding subtest values for the spring grade four test administration are 83.28, 81.33, 82.90 and 75.78. The mean number of 97 items answered correctly in spring is 80.66 percent. For each of the four SMRT subtests and the total test, it is apparent that results from the spring test administration reflect higher achievement than those obtained from the fall test administration.

In order to assess longitudinal trends, the 921 fall 1986 and the 889 spring 1987 grade four student answer sheets were matched to obtain 709 pairs of scores from students who attended both fall and spring test administrations. In other words, these data are longitudinal in the sense that each of the 709 students contributed both fall and spring scores. To accomplish the computerized match of fall and spring test scores, unique nine-digit student identification numbers were used. Subsequently, the accuracy of all matched pairs was verified by visual examination of students' names and dates of birth.

Grade four means and medians based upon longitudinally matched individual student scores are presented in Tables 15 and 16, respectively. On the whole, these grade four longitudinal results reported in Tables 15 and 16 are similar to the cross-sectional data reported in Tables 11 and 12 with two exceptions.

First, the P.S. 290 grade four mean (84.37) obtained in spring based upon longitudinal data was notably higher than the corresponding mean (78.43) based upon cross-sectional data. This difference can be attributed to the inclusion in spring 1987, of relatively low scores from 51 students who were not tested in fall 1986 and, hence, whose scores were not included in the longitudinal analyses.

Second, the P.S. 90 longitudinally based mean (69.79) obtained in fall was much lower than that obtained in fall (75.61) for cross-sectional data. This is due to the fact that fall 1986 cross-sectional data included scores, which were not included in longitudinal analyses, from 23 relatively high achieving students. Specifically, during spring 1987, 23 high achieving P.S. 90 fourth grade students were administered an alternate experimental form of SMRT. The results of the experimental test form are not included in this discussion. Consequently, the P.S. 90 fourth grade longitudinal results for fall 1986 and spring 1987, and cross-sectional results for spring 1987 are somewhat lower (see, in particular, PS 90 results in Tables 11, 12, 15 and 16) than they might have been if results from the high achieving students were included with the results of the other students.

Longitudinally matched subtest and total test grade four results for fall and spring are depicted for all three districts combined ("grade four citywide") in Figure 9, and for each of the three participating Community School Districts in Figures 10, 11 and 12. Subsequently, Figures 13 through 15 depict the increase of longitudinally matched grade four scores from fall to spring. The discrete points shown in these figures have been connected by lines to emphasize the increase. These lines represent an interpolation between the fall and spring scores rather than actual test scores. Furthermore, it is not assumed that increase in achievement is linear.

On the whole, the grade four increases were smaller than those for grade three. Grade three increases for the three participating school districts ranged from 9.51 to 13.30 points. In contrast, the corresponding grade four increases ranged from 5.57 to 9.11 points. These differences may be due, in part, to a possible "ceiling effect." A "ceiling effect" is observed when most of the scores are bunched near the top of the scale. Consequently, there is relatively little opportunity for improvement in subsequent test administrations. In this instance, there is more room for improvement in grade three than there is in grade four.

The difference between fall and spring results for the total test and subtests, respectively, is depicted for all three districts combined ("grade four citywide") in Figures 13 and 14. As with grade three, all subtest scores improved from fall to spring. Students seem to have the greatest difficulty with the reasoning comprehension subtest. A comparison of the school districts (see Figure 15) reveals that both fall and spring results from all three school districts were similar. It is noted that this was not the case for grade three results (see Figure 8) where school district 17 results were relatively lower than those for the other two districts. However, for both grades three and four, all three school districts showed definite improvement from the fall to the spring test administration.

TABLE 11

FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST MEANS FOR GRADE FOUR

Community School District (CSD) and Public School (PS)	Fall 1986			Spring 1987		
	Number Of Students	Mean	Standard Deviation	Number Of Students	Mean	Standard Deviation
CSD 17	487	75.61	14.93	461	81.32	13.87
PS 191	117	76.83	13.10	119	82.43	12.52
PS 289	143	76.62	14.38	131	82.18	13.63
PS 398	227	74.35	16.08	211	80.16	14.68
CSD 19	230	73.22	16.33	263	80.56	14.85
PS 213	108	71.74	17.51	107	81.03	14.98
PS 290	71	74.59	16.04	105	78.43	16.34
PS 328	51	74.45	13.97	51	83.94	10.16
CSD 21	204	77.60	14.35	165	81.91	12.41
PS 90	74	75.61	13.83	44	78.98	13.29
PS 212	75	78.77	14.77	68	83.63	12.22
PS 329	55	78.69	14.40	53	82.13	11.67
TOTAL	921	75.46	15.22	889	81.20	13.91

TABLE 12

FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST MEDIANS FOR GRADE FOUR

Community School District (CSD) and Public School (PS)	Fall 1986			Spring 1987		
	Number Of Students	Median	Semi Inter-Quartile Range	Number Of Students	Median	Semi Inter-Quartile Range
CSD 17	487	79.00	8.50	461	85.00	8.00
PS 191	117	79.00	7.25	119	86.00	8.00
PS 289	143	79.00	8.50	131	85.00	8.00
PS 398	227	78.00	9.00	211	84.00	8.50
CSD 19	230	75.00	11.63	263	85.00	9.00
PS 213	108	75.00	12.38	107	85.00	8.50
PS 290	71	74.00	13.00	105	82.00	10.50
PS 328	51	76.00	10.00	51	86.00	7.50
CSD 21	204	81.00	9.38	165	84.00	7.50
PS 90	74	78.00	10.50	44	81.50	7.38
PS 212	75	83.00	8.50	68	88.00	8.38
PS 329	55	81.00	6.50	53	86.00	6.25
TOTAL	921	78.00	9.50	889	85.00	8.00

TABLE 13

FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST SUBTEST MEANS FOR GRADE FOUR

SMRT Subtests	Number of Items*	Fall 1986		Spring 1987	
		Mean	Standard Deviation	Mean	Standard Deviation
Word Attack	18	14.19	3.14	14.99	2.86
Word Meaning	21	15.91	4.07	17.08	3.64
Literal Comprehension	31	23.71	5.38	25.70	4.86
Reasoning Comprehension	27	18.70	4.50	20.46	4.40
Total	97	72.51	15.16	78.24	13.87

* The three word recognition items have been eliminated from these analyses.

TABLE 14

FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST SUBTEST MEDIANS FOR GRADE FOUR

SMRT Subtests	Number of Items*	Fall 1986		Spring 1987	
		Median	Semi Inter-Quartile Range	Median	Semi Inter-Quartile Range
Word Attack	18	15.00	2.50	16.00	2.00
Word Meaning	21	17.00	2.50	18.00	2.00
Literal Comprehension	31	25.00	3.50	27.00	2.75
Reasoning Comprehension	27	19.00	3.00	21.00	3.00
Total	97	75.00	8.50	82.00	8.00

* The three word recognition items have been eliminated from these analyses.

TABLE 15

LONGITUDINAL FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST MEANS FOR GRADE FOUR

Community School District (CSD) and Public School (PS)	Number Of Students	Fall 1986		Spring 1987	
		Mean	Standard Deviation	Mean	Standard Deviation
CSD 17	382	76.06	14.87	82.98	12.23
PS 191	91	77.45	13.28	84.30	10.59
PS 289	109	77.00	14.04	84.45	10.44
PS 398	182	74.79	16.02	81.43	13.78
CSD 19	182	74.22	16.07	83.33	12.23
PS 213	88	71.90	17.40	82.24	13.33
PS 290	54	75.78	16.08	84.37	11.62
PS 328	40	77.23	12.14	84.33	10.44
CSD 21	145	76.29	13.82	81.86	12.52
PS 90	43	69.79	11.77	78.58	13.19
PS 212	59	78.44	14.85	83.68	12.54
PS 329	43	79.84	12.23	82.65	11.39
TOTAL	709	75.63	14.98	82.84	12.28

TABLE 16

LONGITUDINAL FALL 1986 AND SPRING 1987 SCHOOL MASTERY OF READING TEST MEDIANS FOR GRADE FOUR

Community School District (CSD) and Public School (PS)	Number Of Students	Fall 1986		Spring 1987	
		Median	Semi Inter-Quartile Range	Median	Semi Inter-Quartile Range
CSD 17	382	79.00	9.00	86.00	7.50
PS 191	91	79.00	7.50	87.00	5.50
PS 289	109	80.00	9.75	87.00	6.75
PS 398	182	79.00	8.50	85.00	8.50
CSD 19	182	77.00	10.63	86.50	7.63
PS 213	88	75.00	13.00	86.00	8.00
PS 290	54	75.00	13.00	87.00	8.13
PS 328	40	79.50	8.75	87.50	7.38
CSD 21	145	79.00	9.50	84.00	7.50
PS 90	43	69.00	9.00	81.00	7.50
PS 212	59	83.00	12.00	89.00	8.50
PS 329	43	82.00	5.00	86.00	4.50
TOTAL	709	79.00	9.50	86.00	7.50

Figure 9

Longitudinal Fall 1986 and Spring 1987 Grade 4 Citywide Results

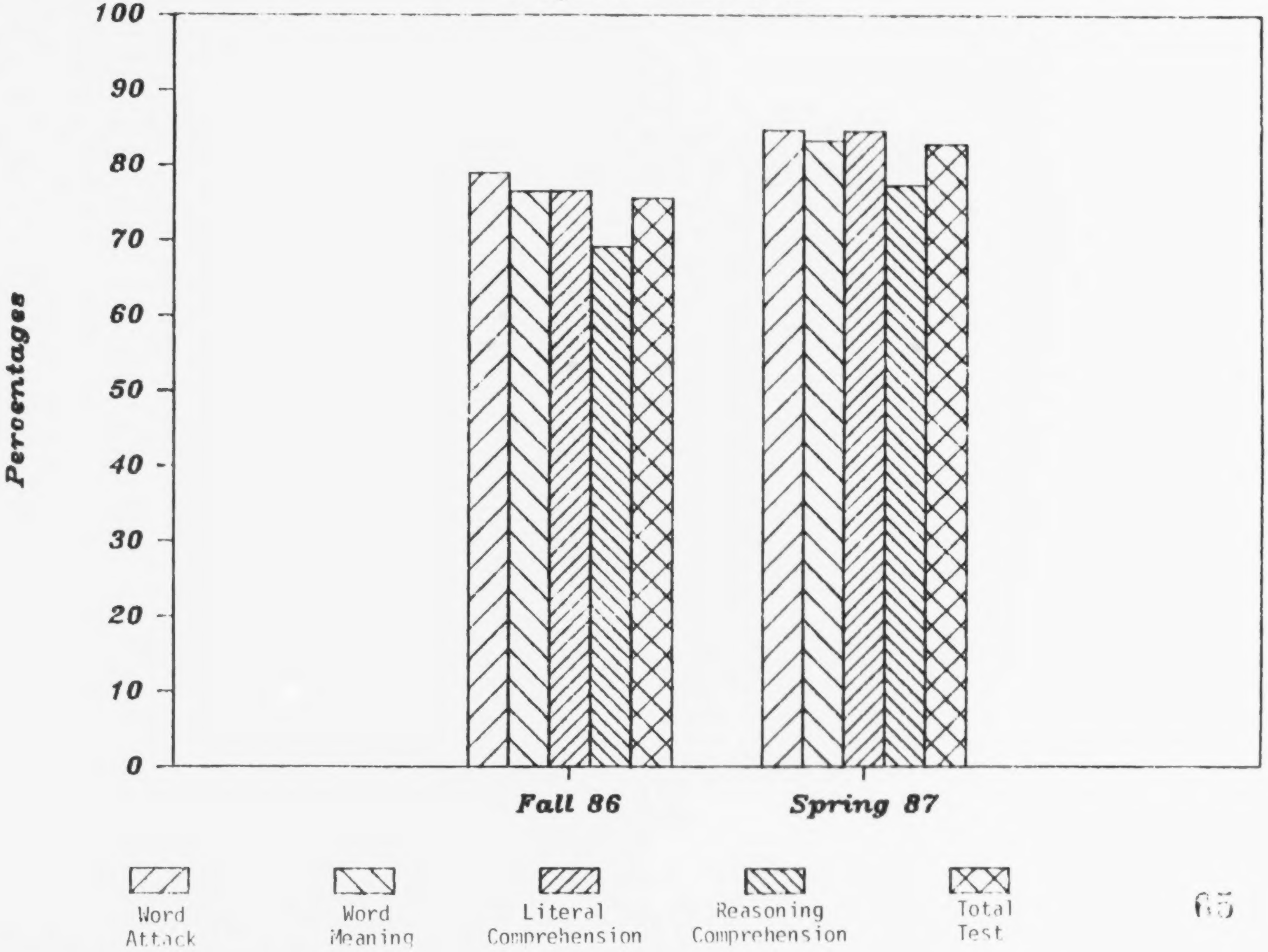


Figure 10

Longitudinal Fall 1986 and Spring 1987 Grade 4 District 17 Results

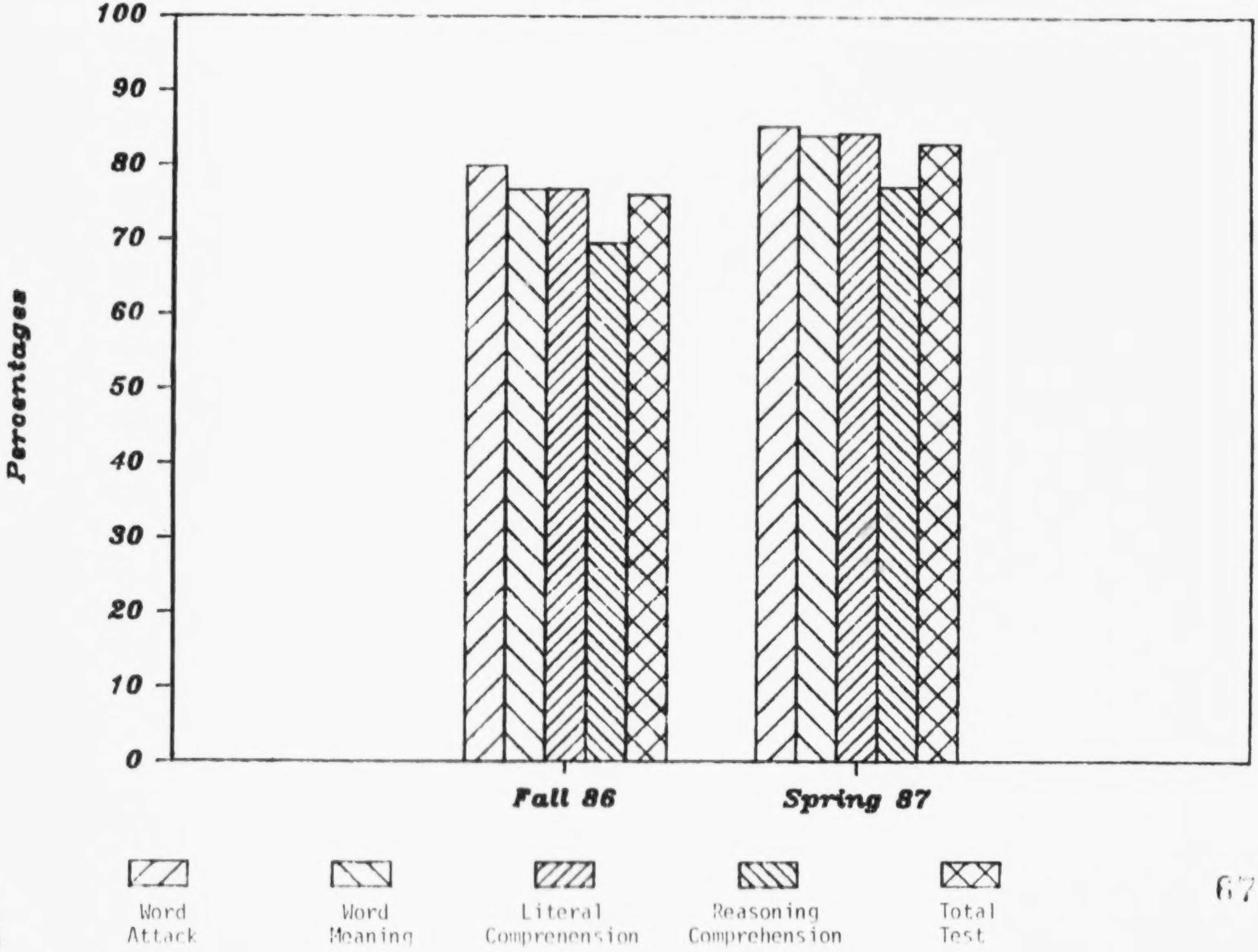


Figure 11

Longitudinal Fall 1986 and Spring 1987 Grade 4 District 19 Results

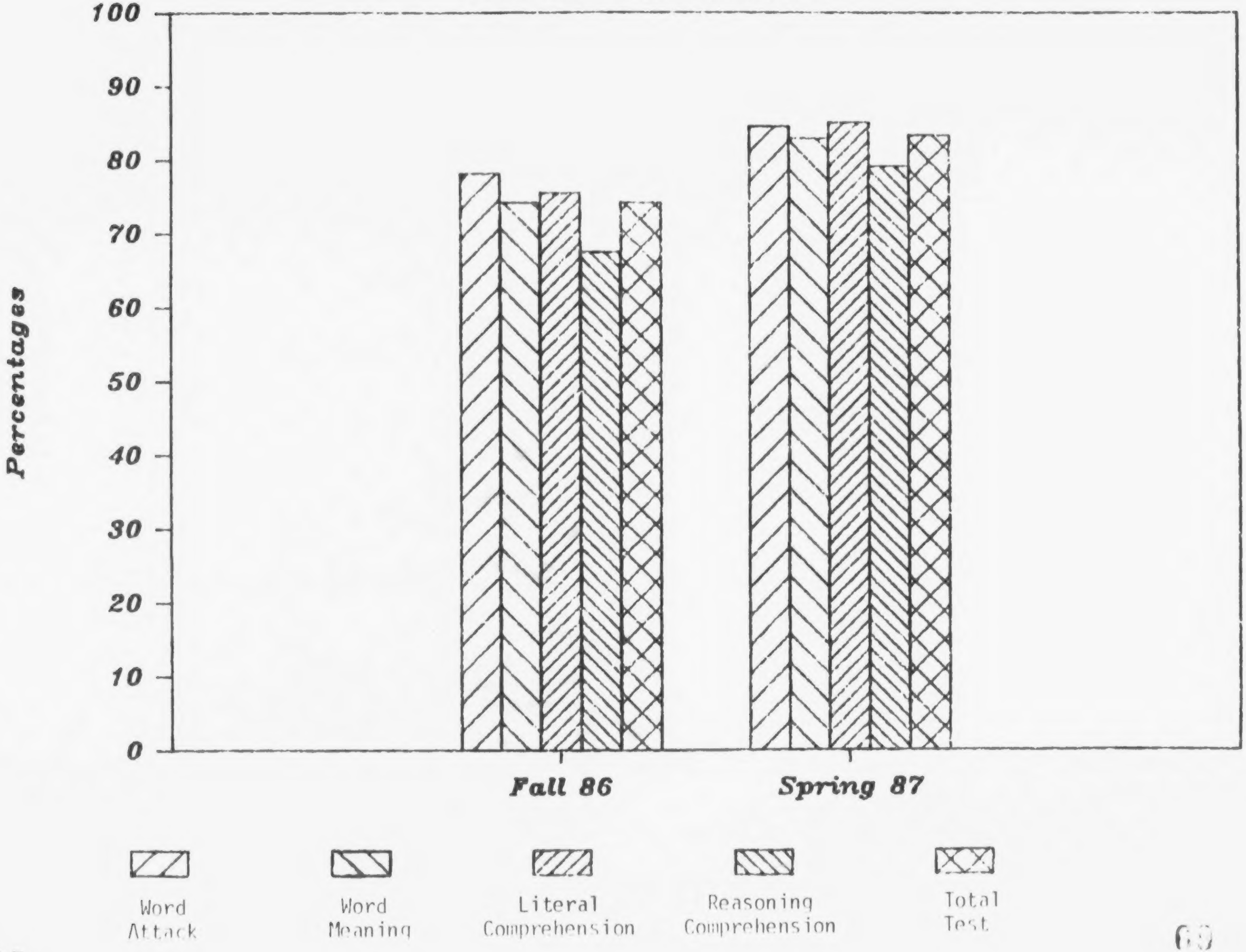


Figure 12

Longitudinal Fall 1986 and Spring 1987 Grade 4 District 21 Results

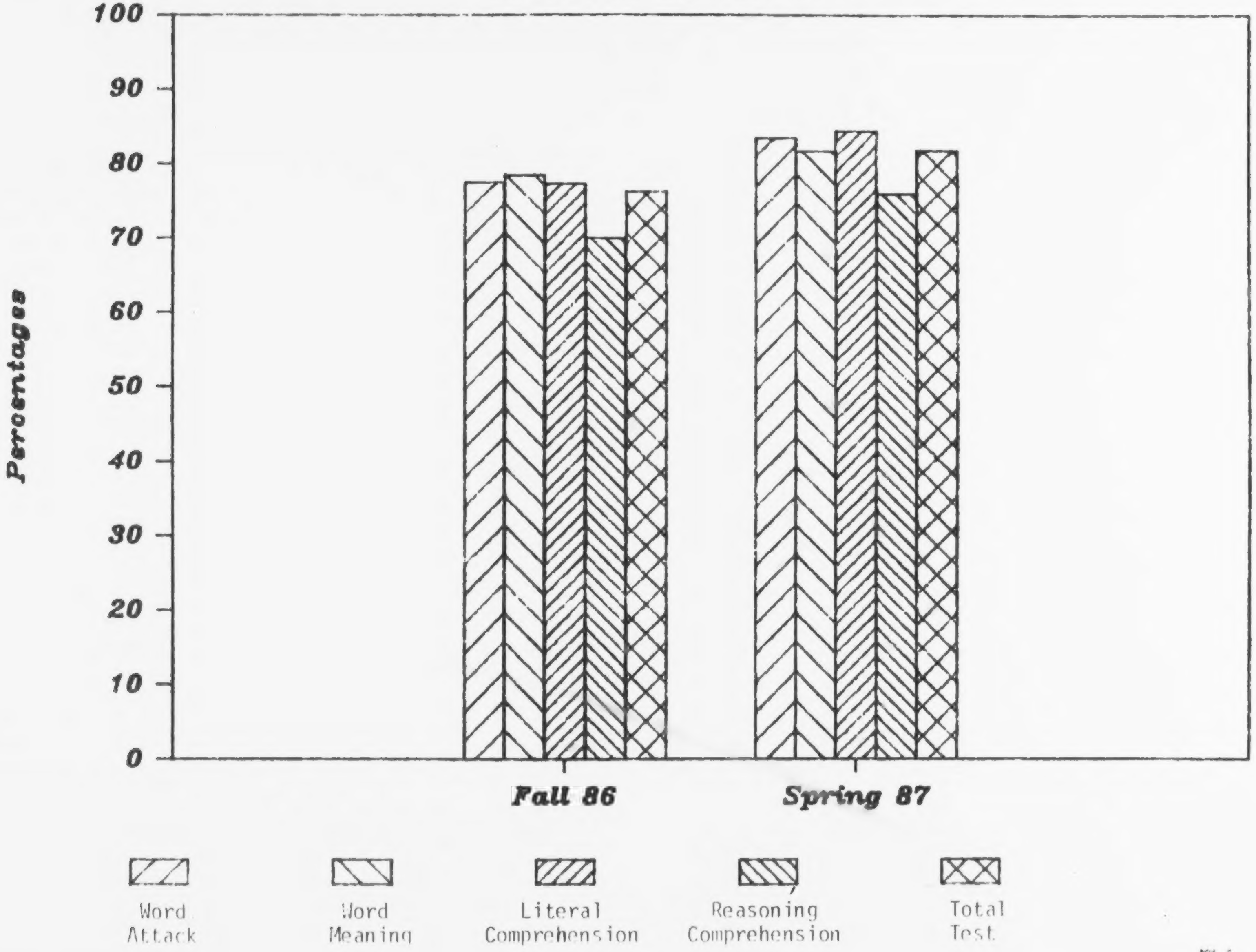


Figure 13

Longitudinal Fall 1986 and Spring 1987 Grade 4
Citywide Total Test Results

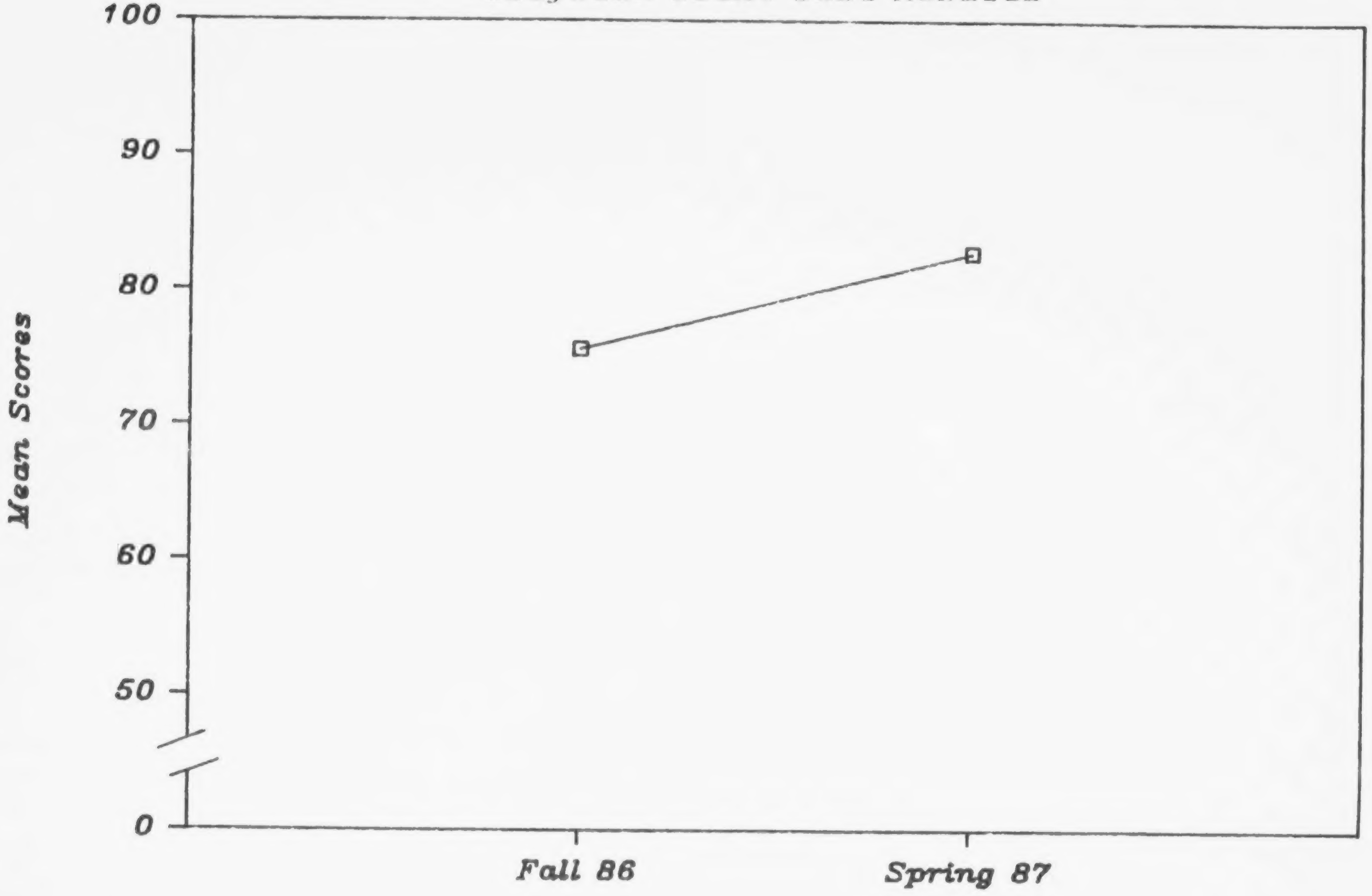


Figure 14

Longitudinal Fall 1986 and Spring 1987 Grade 4
Citywide Subtest Results

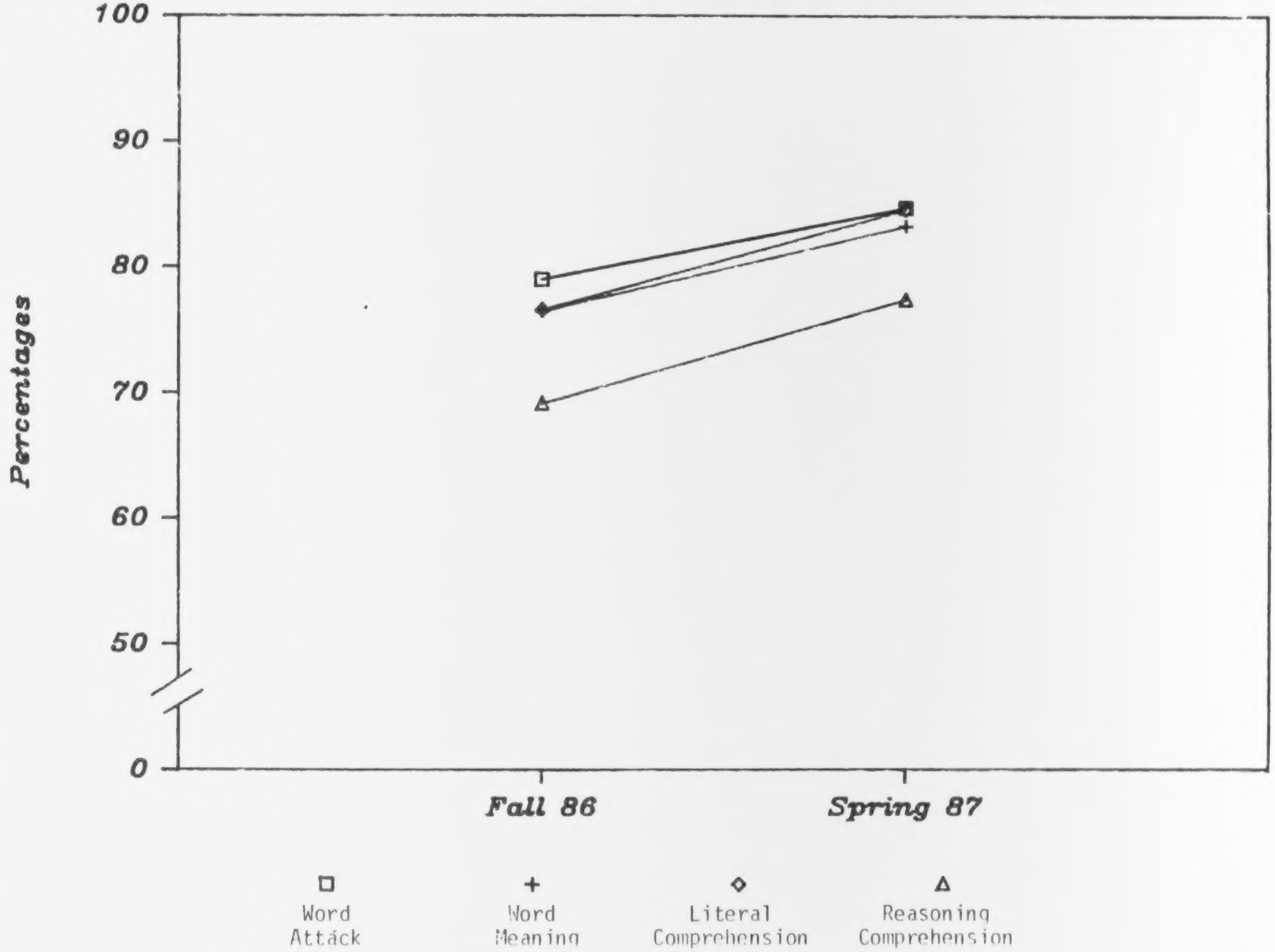
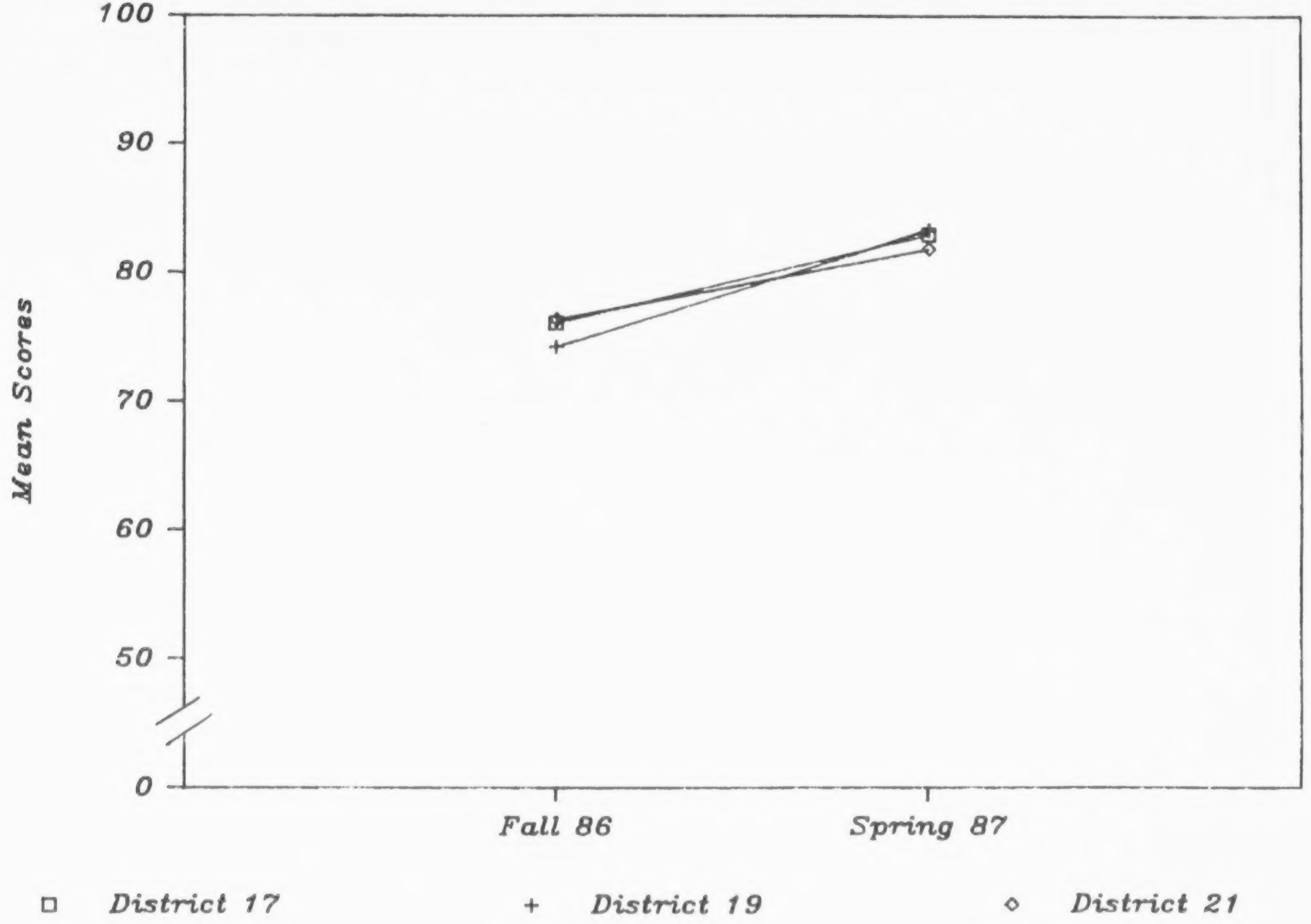


Figure 15

Longitudinal Fall 1986 and Spring 1987 Grade 4 Total Test Results For Three Community School Districts



Spring 1987 Results: A Comparison of Grade 3 and Grade 4

For SMRT to be considered a valid measure of reading achievement, it is necessary to demonstrate that fourth grade students obtain higher scores on the test than third grade students for each of the fall and spring test administrations. Kippel and Forehand (1986, esp. pp. 9-22) demonstrated that grade four SMRT subtest and total test scores resulting from a fall administration were consistently higher than corresponding grade three scores in all three participating Community School Districts.

The following section demonstrates that grade four subtest and total test scores resulting from a spring administration also were consistently higher than corresponding grade three results. It should be noted that Tables 17 through 20 and Figures 18 through 22 repeat data presented in earlier forms in this report. These new tables have been generated for ease of reference.

Examination of Tables 17 through 20 reveals that spring 1987 grade four SMRT subtest and total test scores are consistently higher than those for grade three in all three participating Community School Districts. Total test score means and medians are presented in Tables 17 and 18, respectively, for each of the nine participating schools. Means and medians, respectively, for each of the four SMRT subtests, in addition to the total test, are presented in Tables 19 and 20.

For all three districts combined, the spring 1987 total test mean score for grade three is 74.66 with a standard deviation of 15.89 (see Table 17), and the median is 78.00 with a semi-interquartile range of 10.88 (see Table 18). The spring 1987 total test mean score for grade four is 81.20 with a standard deviation of 13.91 (see Table 17), and the median and semi-interquartile range are 85.00 and 8.00 (see Table 18), respectively. From these tables it is evident that, in all but one of the nine participating schools, grade four students achieve higher SMRT scores than grade three students. The exception was PS 90 (see, in particular, Tables 17 and 18) where the loss of 23 high achieving grade four students in spring 1987, (as explained earlier) caused the spring 1987 grade four results to be lower than it might have been.

For grade three, the mean number of items (presented in Table 19) answered correctly for the word attack, word meaning, literal comprehension and reasoning comprehension subtests represent 77.17, 73.95, 76.19 and 69.11 percent, respectively, of the items on each subtest. The mean number of the 97 items (i.e., the three word recognition items were not included in these analyses) answered correctly for grade three represents 73.92 percent. The corresponding values for grade four are 83.28, 81.33, 82.90 and 75.78. The mean number of 97 items answered correctly for grade four is 78.24 percent. For each of the four SMRT subtests, it is apparent that grade four students achieve higher scores than grade three students. Furthermore, both third and fourth grade students obtained the highest percentage of items correct on the word attack subtest and lowest on the reasoning comprehension subtest (see, especially, Figures 16 and 21). This is consistent with curriculum and instruction emphasis.

Both subtest and total test results for grades three and four are depicted for all three districts combined ("citywide") in Figure 16, and for each of the three participating Community School Districts in Figures 17, 18, and 19. Subsequently, Figures 20 through 22 depict the difference between spring 1987 grade three and grade four SMRT scores. The discrete points shown in Figures 20 through 22 represent data obtained from different grade three and grade four students. The dotted lines between the discrete points have been added to illustrate the differences. The dotted lines do not represent test scores.

The difference between grade three and four total test and subtest results is depicted for all three districts combined ("citywide") in Figures 20 and 21, respectively. All four subtest scores show an approximately equal increase between grades. Both grades obtained highest scores on the word attack subtest and lowest on reasoning comprehension. For both grades, furthermore, there is a notable difference between the performance on the reasoning comprehension subtest and the other three subtests, which are quite close to each other.

The difference between grade three and four total test results for each of the three participating Community School Districts is depicted in Figure 22. It is evident from this figure that all districts show a gain in scores between grades. It is apparent, also, that school district 21 is the highest scoring district.

TABLE 17

SPRING 1987 SCHOOL MASTERY OF READING TEST
MEANS FOR GRADES THREE AND FOUR

Community School District (CSD) and Public School (PS)	Grade Three			Grade Four		
	Number Of Students	Mean	Standard Deviation	Number Of Students	Mean	Standard Deviation
CSD 17	489	73.83	15.98	461	81.32	13.87
PS 191	128	75.61	14.67	119	82.43	12.52
PS 289	173	76.18	15.06	131	82.18	13.63
PS 398	188	70.45	17.11	211	80.16	14.68
CSD 19	280	73.84	16.77	263	80.56	14.85
PS 213	116	74.98	15.99	107	81.03	14.98
PS 290	115	71.97	18.57	105	78.43	16.34
PS 328	49	75.51	13.71	51	83.94	10.16
CSD 21	235	77.35	14.31	165	81.91	12.41
PS 90	65	79.85	13.06	44	78.98	13.29
PS 212	95	76.69	15.02	68	83.63	12.22
PS 329	75	76.03	14.34	53	82.13	11.67
TOTAL	1004	74.66	15.89	889	81.20	13.91

TABLE 18

SPRING 1987 SCHOOL MASTERY OF READING TEST
MEDIANS FOR GRADES THREE AND FOUR

Community School District (CSD) and Public School (PS)	Grade Three			Grade Four		
	Number Of Students	Median	Semi Inter-Quartile Range	Number Of Students	Median	Semi Inter-Quartile Range
CSD 17	489	77.00	11.00	461	85.00	8.00
PS 191	128	78.00	8.88	119	86.00	8.00
PS 289	173	78.00	9.25	131	85.00	8.00
PS 398	188	72.00	13.50	211	84.00	8.50
CSD 19	280	79.00	11.88	263	85.00	9.00
PS 213	116	79.00	9.88	107	85.00	8.50
PS 290	115	78.00	13.50	105	82.00	10.50
PS 328	49	77.00	7.75	51	86.00	7.50
CSD 21	235	80.00	9.00	165	84.00	7.50
PS 90	65	82.00	7.50	44	81.50	7.38
PS 212	95	79.00	8.50	68	88.00	8.38
PS 329	75	80.00	10.00	53	86.00	6.25
TOTAL	1004	78.00	10.88	889	85.00	8.00

TABLE 19

SPRING 1987 SCHOOL MASTERY OF READING TEST SUBTEST MEANS FOR GRADES THREE AND FOUR

SMRT Subtests	Number of Items*	Grade Three		Grade Four	
		Mean	Standard Deviation	Mean	Standard Deviation
Word Attack	18	13.89	3.26	14.99	2.86
Word Meaning	21	15.53	4.12	17.08	3.64
Literal Comprehension	31	23.62	5.63	25.70	4.86
Reasoning Comprehension	27	18.66	4.74	20.46	4.40
Total	97	71.70	15.84	78.24	13.87

* The three word recognition items have been eliminated from these analyses.

TABLE 20

SPRING 1987 SCHOOL MASTERY OF READING TEST SUBTEST MEDIANS FOR GRADES THREE AND FOUR

SMRT Subtests	Number of Items*	Grade Three		Grade Four	
		Median	Semi Inter-Quartile Range	Median	Semi Inter-Quartile Range
Word Attack	18	15.00	2.50	16.00	2.00
Word Meaning	21	16.00	3.00	18.00	2.00
Literal Comprehension	31	25.00	4.00	27.00	2.75
Reasoning Comprehension	27	19.00	3.00	21.00	3.00
Total	97	75.00	10.50	82.00	8.00

* The three word recognition items have been eliminated from these analyses.

Figure 16

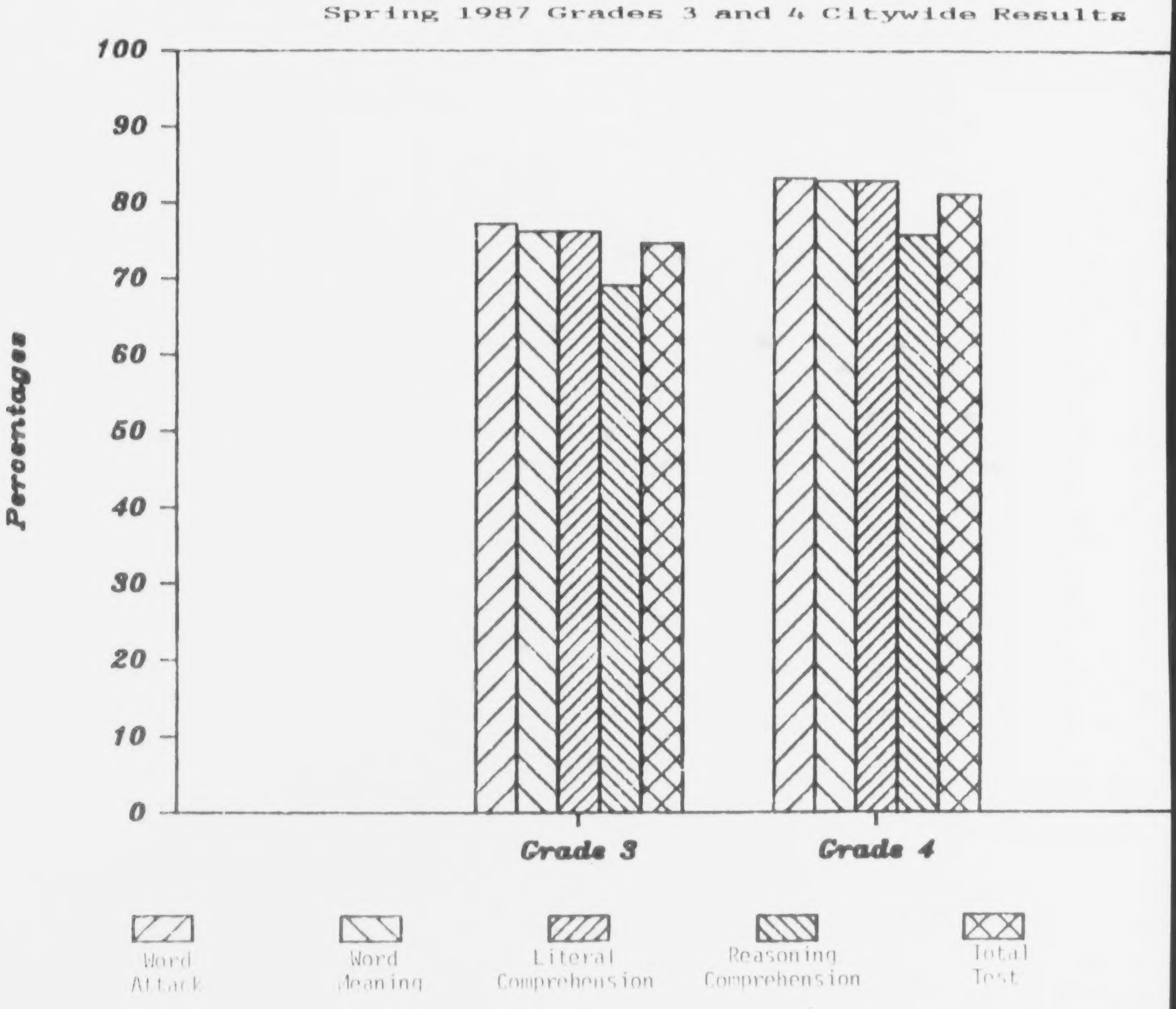

Spring 1987 Grades 3 and 4 Citywide Results
Percentages
100
90
80
70
60
50
40
30
20
10
0
Grade 3
Grade 4
Word Attack
Word Meaning
Literal Comprehension
Reasoning Comprehension
Total Test

Figure 17

Spring 1987 Grades 3 and 4 District 17 Results

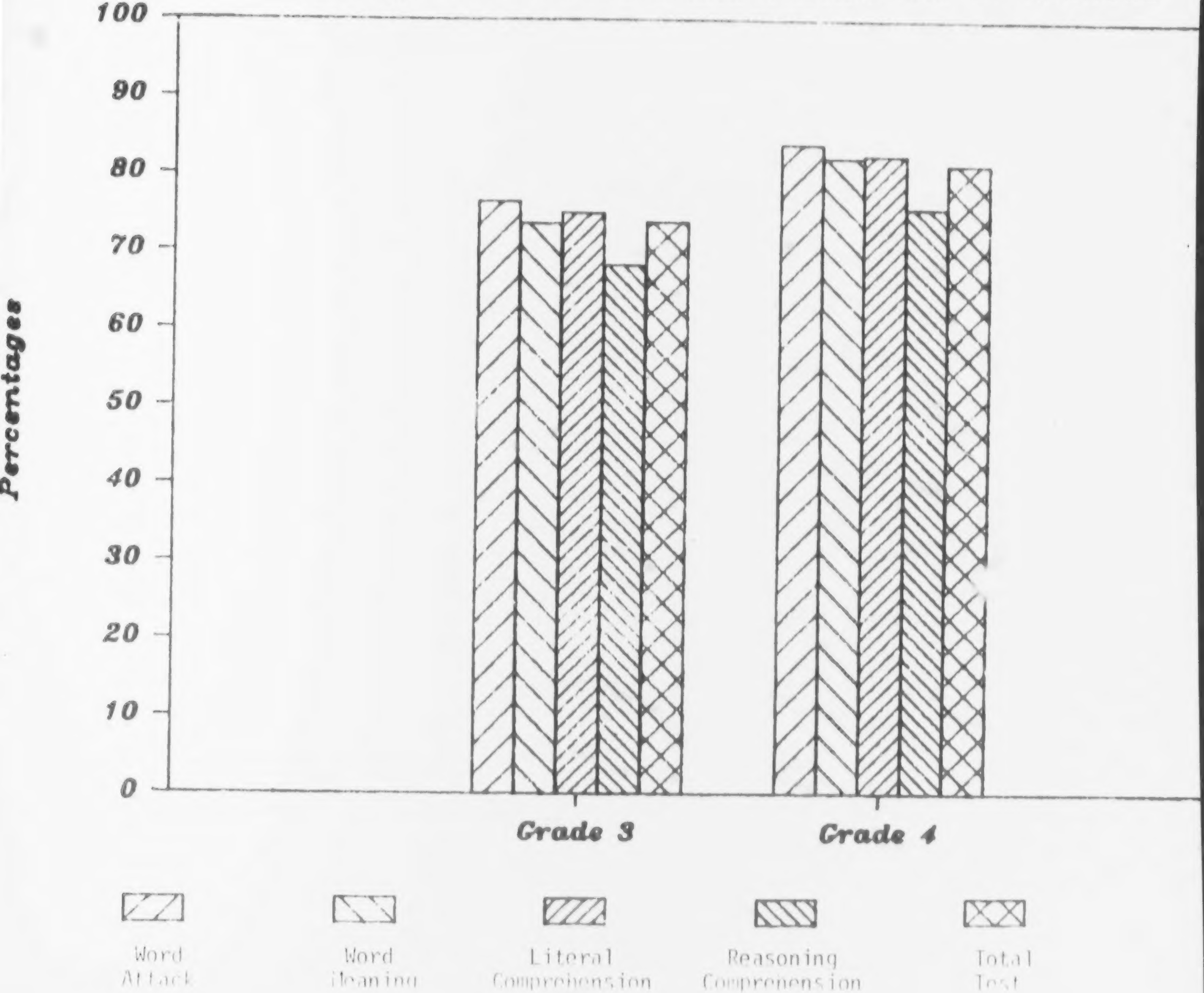


Figure 18

Spring 1987 Grades 3 and 4 District 19 Results

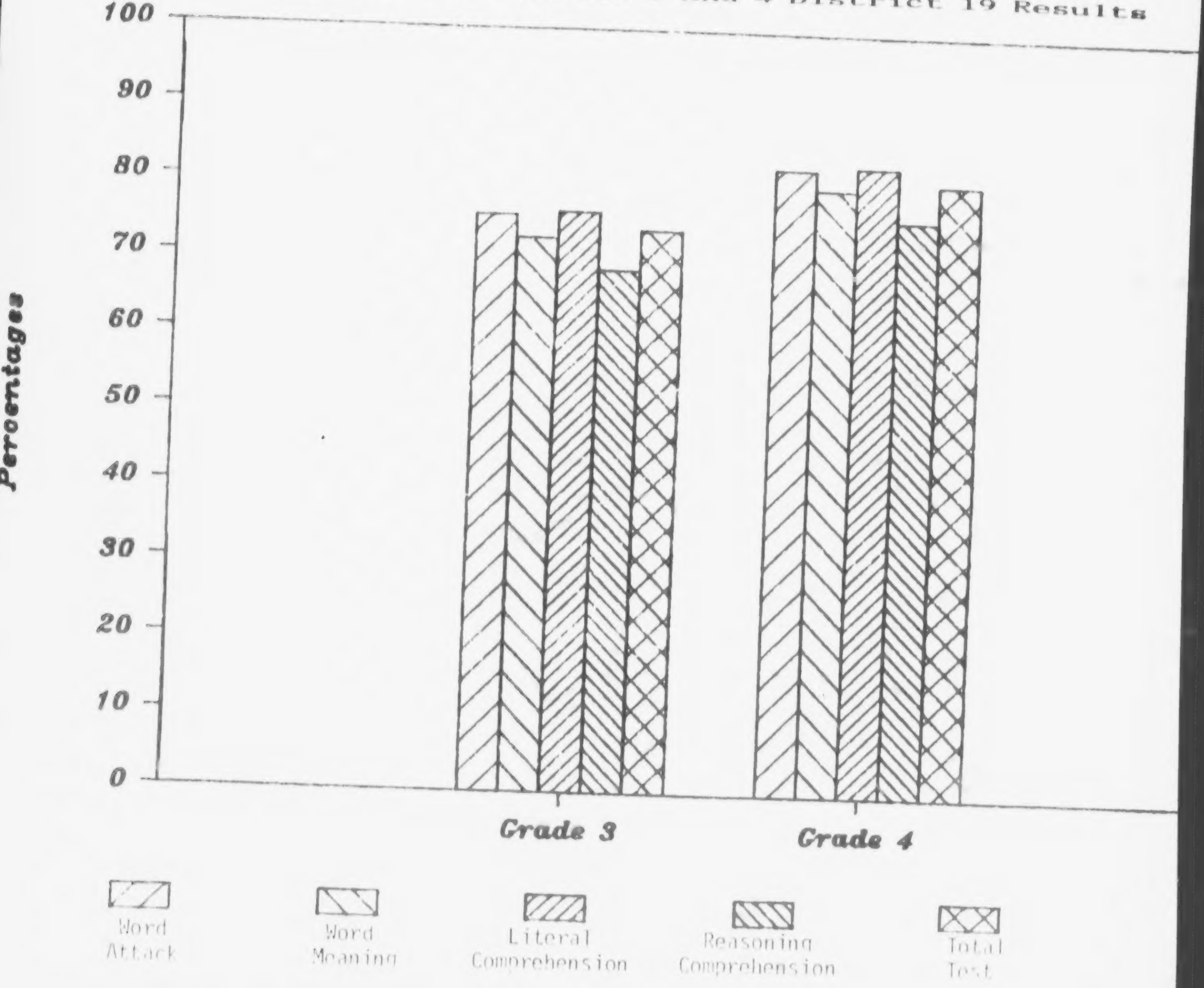


Figure 19

Spring 1987 Grades 3 and 4 District 21 Results

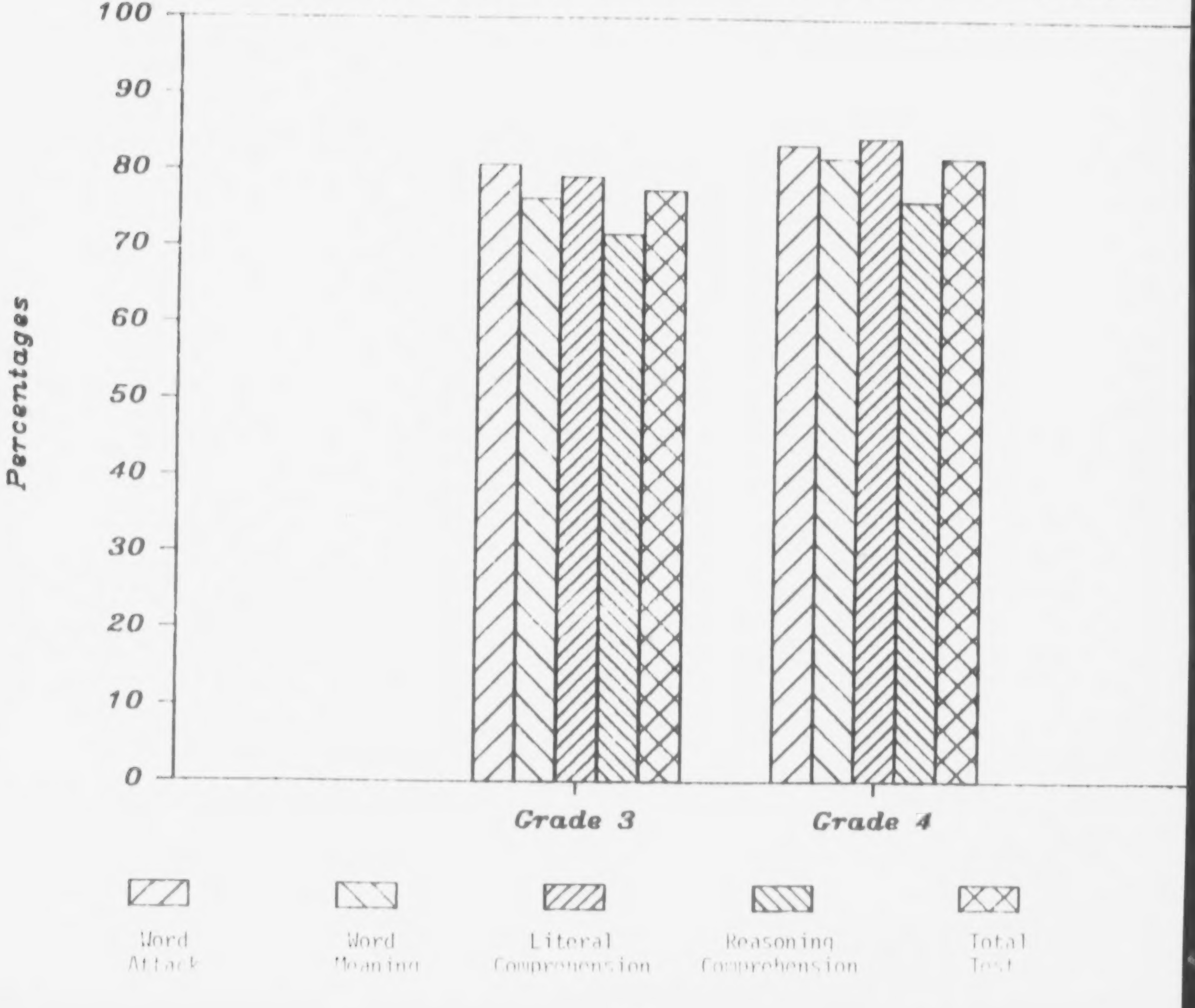


Figure 20

Spring 1987 Grades 3 and 4 Citywide Total Test Results

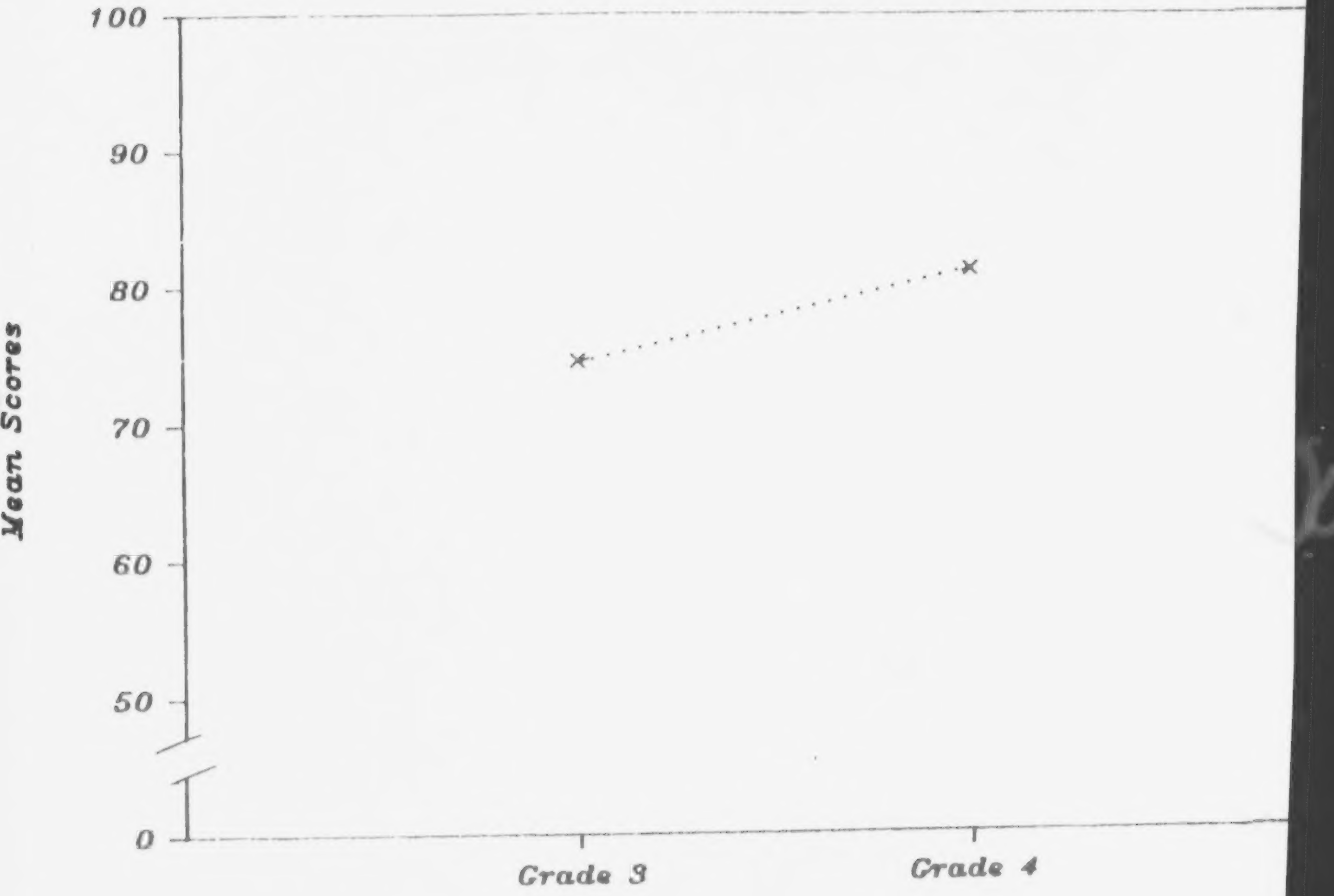


Figure 21

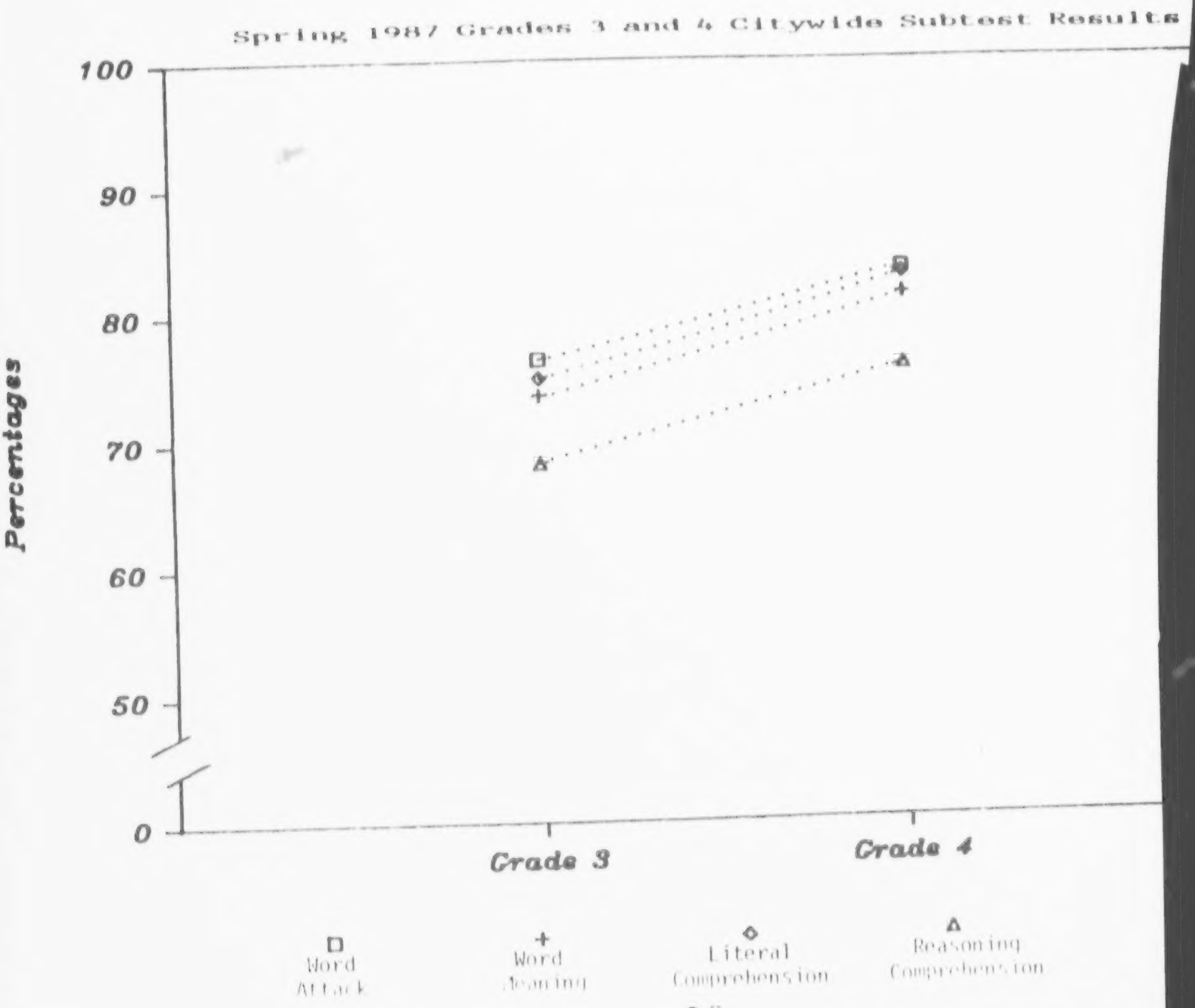
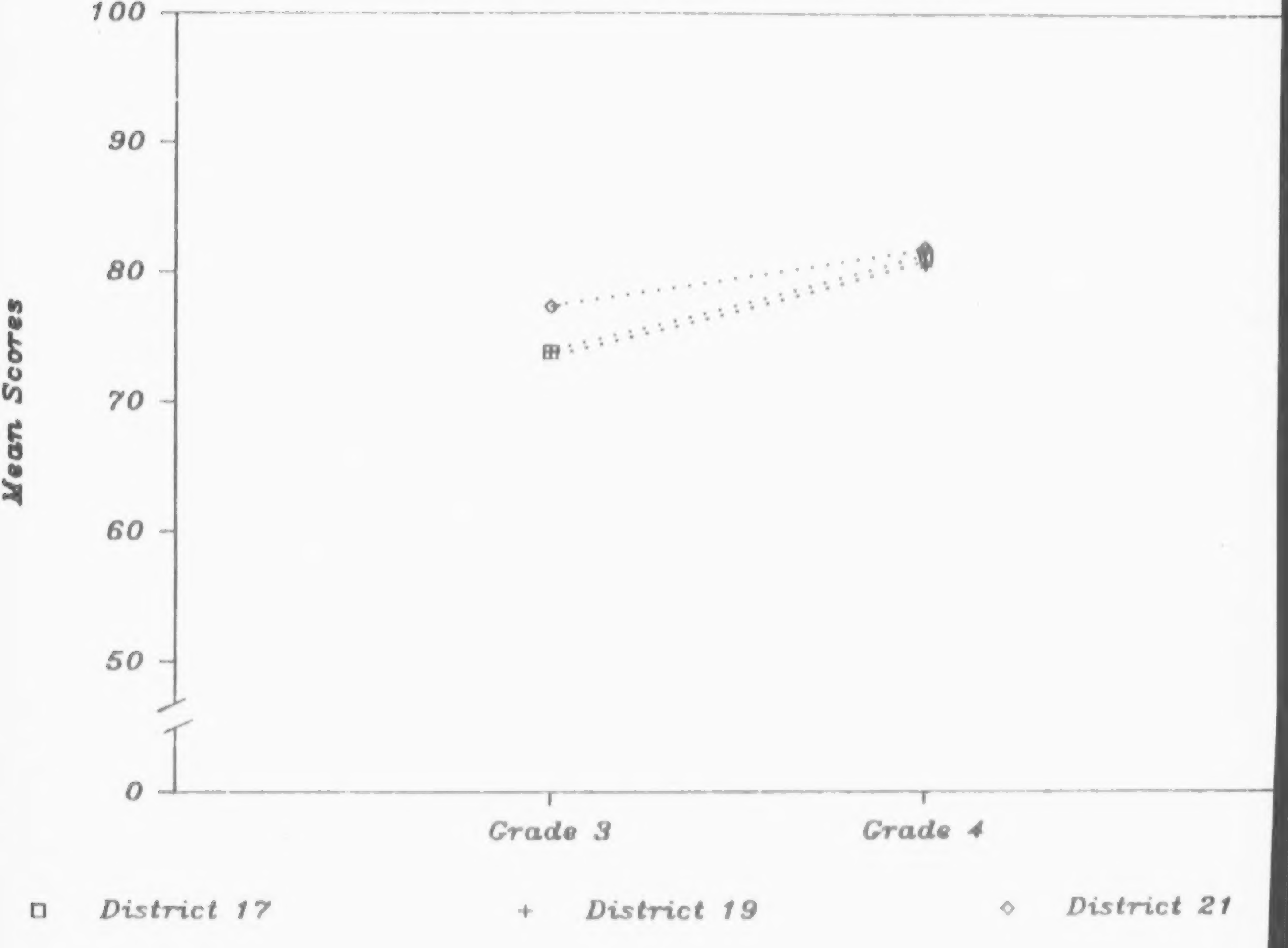

Spring 1987 Grades 3 and 4 Citywide Subtest Results
Percentages
100
90
80
70
60
50
0
Grade 3
Grade 4
□ Word Attack
+ Word Meaning
◇ Literal Comprehension
△ Reasoning Comprehension

Figure 22

Spring 1987 Grades 3 and 4 Total Test Results For Three Community School Districts



VII. RELIABILITY OF THE SCHOOL MASTERY OF READING TEST (SMRT) FOR GRADES THREE AND FOUR

Indices of reliability provide an indication of the extent to which a particular measurement is consistent and reproducible (Thorndike & Hagen, 1977). In other words, reliability refers to the necessity for dependability in measurement (Kerlinger, 1973). Reliability implies stability, consistency, predictability and accuracy. In more technical terms, reliability is the proportion of true variance in obtained test scores (see, for explanation, Guilford, 1954).

Coefficient alpha is the basic formula for determining the reliability based on obtained internal consistency (Nunnally, 1978). Also, it is the expected correlation of one test with an alternative form of the test of the same length, when the two tests purport to measure the same thing.

The grade three and grade four reliability estimates, resulting from both fall and spring administration of SMRT, are presented in Tables 21 and 22, respectively. These data provide support for the contention that SMRT can be used reliably.

Table 21

Reliability of the School Mastery of Reading Test
Fall 1986 and Spring 1987
For Grade Three

Aggregate of Test Items	Number of Items*	Cronbach's Alpha Fall 1986	Cronbach's Alpha Spring 1987
Total Test	97	.9510	.9444
Part One	47	.9032	.8908
Part Two	50	.9152	.9069
Word Attack Subtest	18	.7830	.7708
Word Meaning Subtest	21	.8455	.8212
Literal Comprehension Subtest	31	.8837	.8689
Reasoning Comprehension Subtest	27	.8226	.8208

*The three word recognition items were eliminated from these analyses. Total test reliability, therefore, was based upon 97 rather than 100 items.

Table 22

Reliability of the School Mastery of Reading Test
Fall 1986 and Spring 1987
For Grade Four

Aggregate of Test Items	Number of Items*	Cronbach's Alpha	
		Fall 1986	Spring 1987
Total Test	97	.9351	.9355
Part One	47	.8753	.8812
Part Two	50	.8902	.8946
Word Attack Subtest	18	.7469	.7490
Word Meaning Subtest	21	.8016	.7854
Literal Comprehension Subtest	31	.8513	.8496
Reasoning Comprehension Subtest	27	.7889	.8139

*The three word recognition items were eliminated from these analyses. Total test reliability, therefore, was based upon 97 rather than 100 items.

VIII. DEVELOPMENT OF SUBTESTS FOR THE SCHOOL MASTERY OF READING TEST (SMRT)

Test items were categorized by subtest based upon the professional opinions of several curriculum, reading, research and teaching specialists. Subtests were developed using the definitions provided earlier in Table 2. The following presents correlational evidence relating to the validity of the SMRT subtests.

The School Mastery of Reading Test (SMRT) is comprised of two sections (i.e., Parts I and II), each containing 50 items. The five following types of items are included: 1) word recognition (3 items), 2) word attack (18 items), 3) word meaning (21 items), 4) literal comprehension (31 items), and 5) reasoning comprehension (27 items). Each of these subtests serves to measure a particular facet of reading ability.

The three word recognition items were not considered as comprising a subtest. The word recognition items were relatively easy items and were included both for motivational purposes (i.e., to have students begin the test with easy items) and to orient students to the separate answer sheets. Consequently, the three word recognition items were not included in most statistical analyses.

Correlations which depict the relationship between the four SMRT subtests for 889 fourth grade students tested in May 1986 (see, for discussion, Kippel and Forehand, 1986, esp. pp. 4-5) are presented in Table 23. Review of the correlations reveals that the highest correlation (i.e., .764) is between the literal comprehension and reasoning comprehension subtests. The lowest correlation (i.e., .628) is between word attack and reasoning comprehension.

In order to validate and confirm the placement of items within the particular subtests, the use of factor analytic statistical techniques was considered. However, factor analytic procedures do not appear to be appropriate for the development and confirmation of subtests on mastery tests such as SMRT. The factor analytic technique relies on the assumption that test scores are normally distributed, i.e., some scores are high, some are low, and the majority fall somewhere in-between the two extremes. The SMRT, however, is a test measuring reading mastery administered to students at the end of the academic year. As a result, most students obtain relatively high scores because they have mastered fourth grade reading skills. As expected, consequently, the test scores are "negatively skewed" rather than normally distributed. This departure from bivariate normal distribution might confound any results obtained through factor analytic methods. Consequently, an alternative procedure was used to assess the subtests.

Procedures described below involve the calculation of correlation coefficients which provide estimates of relationships between groups of test items. These particular statistical procedures may be thought of in terms of split-half reliability methods. In effect, the internal structure of the test is being examined by determining the extent to which the items relate to each other.

A high correlation among a set of items, for example, suggests that the items may be measuring a common skill. The items involved, then, may be considered a cluster or factor, representative of one of the various dimensions comprising reading performance. This procedure may be used, for example, to validate two alternate or parallel forms of a given subtest. High correlations among items in different subtests might suggest that the items involved should be combined into one rather than different subtests. In a similar manner, low correlations would suggest distinct subtests.

The following analyses were conducted to determine if items were grouped within subtests in an appropriate manner. In some instances, one might correlate one half of a test with the other half, if both parts were considered to be parallel or equal forms. In this instance, however, there is evidence that some students achieved lower scores on Part II compared with Part I in the May 1986 test administration. The cause of this pattern of results is not clear. It may be due, among other reasons, to considerations such as fatigue, relatively stringent time limits, and/or relatively difficult items appearing in Part II compared with Part I. It appears prudent, therefore, not to consider the two parts of SMRT as equal. Consequently, a strategy was implemented which involved rearranging items according to difficulty levels in order to develop an analogue to parallel forms.

In order to accomplish this, the item difficulty was determined for all 97 test questions. The three Word Recognition items were deleted from the original 100 items. Subsequently, items were ranked by difficulty within each of the four subtests. For each of the four subtests, separately, items were matched by difficulty level and redistributed into two modified and parallel halves of each subtest. In effect, each modified subtest in Part I was approximately equal in terms of item difficulty to its corresponding modified subtest in Part II.

Pearson product-moment correlation coefficients were calculated between the modified Part I and modified Part II subtests. For example, the correlation was computed between modified Part I Word Attack and modified Part II Word Attack items. Table 24 presents the correlation of the four modified subtests in Part I with their parallel forms in Part II.

In a similar manner, correlation coefficients were obtained between all the modified subtests within Part I. For example, the correlations were obtained between modified Part I Word Attack and modified Part I Word Meaning, Literal Comprehension and Reasoning Comprehension, respectively. Table 25 presents the correlations between the four different subtests in Part I. Finally, correlation coefficients were obtained between all the modified subtests within Part II. These correlations are reported in Table 26.

As one examines Tables 24, 25 and 26, it becomes evident that the correlations between different subtests are lower than those obtained between the parallel forms within each subtest. For example, the Literal Comprehension subtest in Part I of the SMRT correlates more highly with its Literal Comprehension parallel form in Part II ($r = .71$) than it does with any other of the other subtests in Part II. This finding reinforces the notion that distinct facets of reading performance are assessed by the SMRT subtests.

Further inspection of the data reveals that the correlations in Tables 25 and 26, although lower than those in Table 24, are nonetheless significant at the $p < .01$ level. That is, there is a considerable degree of overlap in different SMRT subtests. It is reasonable to expect some relationship between the different SMRT subtests because each is measuring some aspect of reading performance. Examination of the correlation coefficients in Tables 25 and 26 reveals that the highest correlations in Tables 25 and 26 were $r = .62$ and $r = .67$, respectively. Both of these correlations were obtained between Literal and Reasoning Comprehension subtests. The lowest correlation in Part I was between Word Attack and Reasoning Comprehension ($r = .49$). The lowest correlation in Part II was between Word Attack and Literal Comprehension ($r = .52$).

TABLE 23

Correlations Between the School Mastery of Reading Test (SMRT) Subtests

(n = 889)

SUBTEST	SUBTEST	PEARSON r
Word Attack	Word Meaning	.67 **
Word Attack	Literal Comprehension	.64 **
Word Attack	Reasoning Comprehension	.63 **
Word Meaning	Literal Comprehension	.73 **
Word Meaning	Reasoning Comprehension	.70 **
Literal Comprehension	Reasoning Comprehension	.76 **

**$p < .01$

TABLE 24

Correlations Between Modified Part I and Modified Part II Subtests of the School Mastery of Reading Test (SMRT)

(n = 889)

PART I MODIFIED SUBTEST	PART II MODIFIED SUBTEST	Pearson Product-Moment CORRELATION COEFFICIENT
Word Attack	Word Attack	.61 **
Word Meaning	Word Meaning	.69 **
Literal Comprehension	Literal Comprehension	.71 **
Reasoning Comprehension	Reasoning Comprehension	.69 **

**$p < .01$

TABLE 25

Correlations Between Modified Subtests Within
Part I of the School Mastery of Reading Test (SMRT)
(n = 889)

PART I MODIFIED SUBTEST	PART I MODIFIED SUBTEST	PEARSON r
Word Attack	Word Meaning	.53 **
Word Attack	Literal Comprehension	.53 **
Word Attack	Reasoning Comprehension	.49 **
Word Meaning	Literal Comprehension	.59 **
Word Meaning	Reasoning Comprehension	.56 **
Literal Comprehension	Reasoning Comprehension	.62 **

** $p < .01$

TABLE 26

Correlations Between Modified Subtests Within Part II of the School Mastery of Reading Test (SMRT)

(n = 889)

PART II MODIFIED SUBTEST	PART II MODIFIED SUBTEST	PEARSON r	
Word Attack	Word Meaning	.53	**
Word Attack	Literal Comprehension	.52	**
Word Attack	Reasoning Comprehension	.53	**
Word Meaning	Literal Comprehension	.63	**
Word Meaning	Reasoning Comprehension	.61	**
Literal Comprehension	Reasoning Comprehension	.67	**

** $p < .01$

IX. THE SCHOOL MASTERY OF READING TEST (SMRT) AND NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP) NORMS AND PERFORMANCE STANDARDS

The following demonstrates the manner in which SMRT results may be interpreted with respect to NAEP national norms and performance standards. To some extent NAEP might be a cost-effective source of new test items for SMRT.

National Assessment of Educational Progress

National Assessment of Educational Progress (NAEP) has been developed to measure how effectively 9-, 13- and in-school 17-year-old American students can read (Messick, Beaton, & Lord, 1983). For this purpose, nationally representative samples of students within various demographic subgroups are tested (National Assessment of Education Progress, 1985).

NAEP bases each assessment on a wide range of materials and asks questions requiring use of a variety of reading skills and strategies. Reading selections range from simple sentences expressing a single concept to complex articles about specialized topics in science or social studies. Both items and tests span a wide range of difficulty and are presented in a variety of formats.

Items are reviewed for potential bias before being accepted by NAEP for administration. Specifically, NAEP items are reviewed by educators on the basis of their academic appropriateness, effectiveness, freedom from bias or stereotyping, and sensitivity to racial, ethnic, religious and political groups. After test administration, item response curves are analyzed for potential bias.

The relationship between SMRT and NAEP is being determined. In effect, the current study is designed to improve local school level diagnosis and prescriptions for progress by using NAEP items and norms (See footnote #1). The primary intent is to determine the feasibility of:

1) obtaining norm-referenced interpretations of SMRT results with respect to NAEP national norms

2) demonstrating the extent to which SMRT results relate to NAEP performance standards

3) establishing a cost-effective source of new items by incorporating NAEP items within SMRT

---

[1] In addition, it is noted that a somewhat different potential role for NAEP in assisting the development and implementation of local educational standards has been defined by Messick (1985).

Selection of NAEP Items

In order to achieve these objectives, NAEP items were evaluated with regard to item content, format and general appropriateness for New York City fourth grade students. It was determined, consequently, that some NAEP items could be incorporated within SMRT. This decision was based upon the fact that current elementary school level NAEP items were designed for grade three students and have sufficient range for grade four students. In the recent past, elementary school level NAEP items were designed for grade four students. It is noted that SMRT is designed for relatively low achieving fourth grade students. Furthermore, SMRT is most likely to be administered early in the school year for maximum diagnostic usefulness.

Some NAEP items are so similar in format and content to some SMRT items that, if mixed together, it would be difficult to determine the source of each. At the same time, some SMRT item types are not matched by NAEP items. As indicated in Table 27, the NAEP items appear to be somewhat more difficult than the SMRT items for both third and fourth graders. Specifically, the percentage of NAEP items correct was lower than the corresponding percentage for SMRT items for all three 1986 and 1987 test administrations (i.e., May 1986, October 1986, May 1987).

At our request, the Educational Testing Service (ETS) obtained permission for the use of NAEP items within SMRT. Permission was granted to use NAEP items under "Reasonable constraints". Specifically, it is understood that: 1) NAEP items will not be published or inappropriately disseminated, 2) NAEP items will not be used for pre-test practice or instruction, and 3) appropriate steps will be taken to insure adequate security of NAEP items.

Selection of particular NAEP items for inclusion within SMRT were based upon item scale value, content and format. A total of 16 NAEP comprehension items were selected for testing. These NAEP items were embedded within both Parts I and II of SMRT. The NAEP items are identified in Appendix A.

Scaling NAEP by Item Response Theory

NAEP has applied Item Response Theory (IRT) to define the probability of answering reading exercises correctly as a function of ability level or skill. Specifically, the logistic mathematical function has been used to provide one ability level parameter or measure (i.e., theta) for each individual and three parameters or calibrations for each exercise. The three item parameters reflect discriminating power (a-value), difficulty level (b-value) and likelihood of guessing (c-value) (see, for discussion, Messick, Beaton, & Lord, 1983, pp. 43-55). The item parameters are used for the purposes of equating SMRT ability to national norms based upon NAEP. A full discussion of these procedures follows in the section entitled: "SMRT Results and NAEP norms."

NAEP has developed a scale ranging from 0 through 500 by applying a linear transformation to the ability estimate. Various points on that scale have been provided criterion-referenced interpretations. As explained further in the section entitled "SMRT Results and NAEP Levels of Proficiency," the criterion-referenced interpretation will be validated based upon SMRT, after SMRT has been equated to NAEP.

Demonstrating That SMRT Is Unidimensional

IRT methods are appropriate for unidimensional areas in which the exercises are scored right, wrong or no response. It was necessary, therefore, to test the assumption of unidimensionality of SMRT before IRT methods could be considered appropriate. In particular, it was necessary to demonstrate that SMRT and NAEP items load on the same common scale. If SMRT and NAEP items measure the same underlying reading proficiency variable, items from both tests could be interchanged without disturbing normative and criterion-referenced interpretations of the test scores. It is noted that IRT methods are particularly relevant for facilitating the ultimate goal of tying SMRT into national norms based on NAEP.

To verify the unidimensionality of SMRT, a principal components factor analysis was performed on May 1986 test scores from 889 fourth grade students, to examine the underlying factor structure of the 100 item SMRT. Subsequently, it was demonstrated that 91 of the 100 items comprise a single dimension and meet IRT assumptions. These 91 items included the 16 NAEP items, thus demonstrating that these 16 NAEP items and 75 additional SMRT items load on the common scale. In effect, they measure the same dimension. The remaining nine items reflected relatively low weights on the principal factor and were eliminated from subsequent IRT analyses. It is noted that four (i.e., items 1, 2, 3 and 9) of the nine eliminated items were sample or orientation items not intended for subsequent analyses. The remaining five items (i.e., 4, 24, 63, 91 and 99) will be revised or eliminated, as appropriate, from future editions of SMRT.

Scaling SMRT by Item Response Theory

Next, consistent with methods established and implemented as part of the NAEP program (see, for discussion, Messick, Beaton and Lord, 1983), a three-parameter IRT analysis was conducted on the 91 item unidimensional SMRT. The advantage of IRT methods is to facilitate the equating of NAEP items to the SMRT items. Once these items are equated, it is then possible to estimate the common scale scores of student abilities.

To examine the extent to which the SMRT items measure the same reading proficiency variable as measured by the NAEP test, item characteristics based on the 3-Parameter IRT model were compared for the two tests across the range of reading proficiency (i.e., "theta"). A graphic representation of this study is depicted in Figure 23. In Figure 23, the item characteristics of all 91 SMRT items were summarized by a Test Characteristic Curve (TCC) depicted by a dotted function. The second TCC in Figure 23, depicted by a solid line, summarizes the item characteristics of the 16 NAEP anchor items. It is noted that these 16 NAEP items were included in the TCC for the 91 SMRT items.

Each TCC depicts the probability of mastery (plotted along the vertical axis) for students of any given level of reading proficiency (plotted along the horizontal axis). In other words, each TCC represents the expected level of mastery for the range of possible reading proficiency levels of the students taking the test. Thus, the TCC for the SMRT items can be used to estimate the percentage of SMRT items that a student of any particular reading proficiency would be expected to master. For example, a student with a theta of about -1 would be expected to master approximately 60% of the SMRT items.

To ensure that the SMRT test produces a similar TCC when compared to NAEP items, the TCC for the 91 SMRT items was compared with that based upon the 16 NAEP items only. Inspection of Figure 23 reveals that the respective TCC's for SMRT and NAEP were, indeed, similar across the range of reading proficiency. Since the 16 NAEP items were anchored within the SMRT test, it may be concluded that the SMRT items can be interchanged with NAEP items, and that SMRT test data can be expressed in terms of the normative and/or criterion-referenced interpretations based upon NAEP.

Further inspection of Figure 23 reveals that, while there were no differences between NAEP and SMRT TCC's at the middle range of reading proficiency, the SMRT items yielded slightly higher ability estimates for low ability and high ability students, respectively. This finding will be considered more carefully when NAEP and SMRT are equated. The results of this comparison of TCC's supported the feasibility of expressing SMRT test data in terms of interpretations based on NAEP.

A second set of analyses were performed in order to determine the relative stability of IRT item parameters for the purpose of equating SMRT to NAEP. For these analyses, item "pre-calibrations," which were based upon the NAEP standardization sample (i.e., used by NAEP to promulgate national norms), were compared to "new (i.e., SMRT) estimates," which were derived from the current SMRT administration. In Figure 24, a bivariate plot of the NAEP pre-calibrations and the SMRT estimates is shown for the item difficulty calibrations (b-values). Inspection of Figure 24 reveals that a linear relationship exists for these two sets of estimates based on item difficulty. For equating purposes, a linear trend must exist to ensure that equations based on item difficulties will remain stable over subsequent administrations of the SMRT.

Figure 25 depicts the stability of the item discrimination indices (a-values). Unlike the desirable results based on item difficulties (i.e., see Figure 24), it can be seen that the item discriminations are relatively dispersed around this identity line. The implication of this result is that equating based on item discrimination would be inaccurate from sample to sample. It should be noted that guessing (i.e., c-values) are not ordinarily used for equating purposes.

Based on the findings from these analyses, it was concluded that the equation of SMRT to NAEP should be based on the item difficulty calibrations (i.e., b-values) only. The details of this equation will follow later.

In effect, the validity of calibrating the SMRT items onto the NAEP scale has been demonstrated. Consequently, SMRT results can be interpreted with respect to NAEP national norms and performance standards. Furthermore, SMRT items can be replaced with comparable NAEP items. In order to interchange current SMRT items with previously unused items from the NAEP item pool, the item difficulty or b-value item characteristic parameter would be used. In addition, item content, format and congruence with the original SMRT blueprint must be considered.

SMRT Results and NAEP Norms

The objective is to use the b-value item parameter estimates for the 16 NAEP anchor items which were embedded within SMRT in order to derive SMRT norms and proficiency levels. The b-value estimates that we will "tie into" are those obtained for the 16 NAEP items from the original norming of NAEP. In effect, the overall goal is to establish a common SMRT-NAEP scale with a calibrated item pool.

As noted previously, analytic studies showed that the SMRT can be equated to the NAEP test using item difficulty calibrations (i.e., b-values). Sixteen of the 91 SMRT items were "anchored" -- meaning that these 16 items are actually NAEP items. It is necessary, therefore, to treat these 16 items as an "anchor test" to be used for equating purposes. Since only the item difficulty calibrations will be used for equating, the equating design is referred to as a "one-parameter" or a "Rasch model" horizontal equation. The schematic of this design is depicted in Figure 26.

The mechanics of the equation can be summarized in three steps. First, the b-values of the NAEP items estimated for the NAEP item pool (Item Pool 1) will be compared to the b-values for the same items when administered within the SMRT (Item Pool 2). An identity line should emerge to ensure that the (same) anchor items retained similar item difficulty calibrations for both pools. Items which appear to depart from this assumption will be deleted. The remaining anchor items will be treated as referents to the normative and criterion-related interpretations of the NAEP ability scale.

Second, an equating constant will be estimated in order to translate SMRT item difficulty calibrations in terms of the NAEP item scale. This equating constant will be estimated from a regression analysis of b-values of NAEP items retained from Step 1 (Item Pool 2) on b-values for the same items from their original item pool (Item Pool 1).

Third, the equating constant will be applied to all 91 SMRT items (Item Pool 3) in order to "translate" the SMRT items in terms of the NAEP scale (Item Pool 1). Subsequently, the resulting SMRT item pool can be referenced to normative and criterion-related interpretations based on NAEP. Since IRT facilitates a direct translation from the item difficulty scale to the theta scale, it is possible to estimate a scale score of ability directly from the item difficulty scale. Thus, the equating procedures will facilitate, for example, how SMRT results are interpreted with respect to NAEP norms. Furthermore, new forms and levels of SMRT can be designed which will be based upon New York City curriculum and will yield NAEP norm-referenced interpretations.

SMRT Results and NAEP Levels of Proficiency

Once the SMRT ability scale is equated to that of the NAEP scale, the SMRT results may be interpreted according to NAEP performance standards. These criterion-related interpretations of SMRT in terms of NAEP will also be performed through the equating procedures described previously.

Specifically, Levels of Proficiency have been established by NAEP (see, for explanation, National Assessment of Educational Progress - Report No: 15-R-01, pp.14-30) to describe the kinds of reading tasks that most children, who have reached each level of reading proficiency, are able to do. Each of the five Levels of Proficiency are related to a point on the 0-500 NAEP scale and Table 28 briefly describes each level. According to NAEP, the interaction of the following three factors affects students' reading proficiency: the complexity of the material they are asked to read, their familiarity with the subject matter, and the kinds of questions asked. The many possible interactions among the passage, question, and prior knowledge components are reflected in the NAEP reading proficiency levels.

As indicated earlier, the logistic function has been applied to obtain three item parameters for each of 91 SMRT items, after the additional items were eliminated. Subsequently, the b-value item difficulty calibration of each of the 91 SMRT items was equated (using the translation constant obtained from the 16 NAEP anchor items) with the NAEP ability scale and Levels of Proficiency. This analysis indicates that the SMRT items can be categorized and described as specified in Table 28.

Further Studies

To ensure that the equating procedures produce a reliable translation of SMRT results in terms of NAEP norms and performance standards, follow-up studies will be performed on subsequent SMRT administrations. In addition, an item bank consisting of SMRT items will be created and maintained for future reference. The item bank will be updated and expanded as new SMRT items are tried out. Further, existing item calibrations will be updated to reflect changing characteristics of the curriculum and the student population. In addition, the item bank will permit the assembly of alternate forms of SMRT, with each form equated to NAEP norms and performance standards.

Table 27

Percentage of correct items for the
National Assessment of Educational Progress (NAEP) and the
School Mastery of Reading Test (SMRT)
(97 Items)*

Type of Item (Number of Items)	Grade	Spring 1986	Fall 1986	Spring 1987
NAEP (16 Items)	Four	71.33	63.14	72.25
	Three	(Not Administered)	51.49	63.81
SMRT (81 Items)	Four	81.79	77.05	82.32
	Three	(Not Administered)	65.66	75.91

* The first three (word recognition) items have been eliminated

Table 28
National Assessment of Educational Progress (NAEP)
Levels of Proficiency*

Rudimentary (150) (51 of 91 SMRT Items)

Readers who have acquired rudimentary reading skills and strategies can follow brief written directions. They can also select words, phrases, or sentences to describe a simple picture and can interpret simple written clues to identify a common object. *Performance at this level suggests the ability to carry out simple, discrete reading tasks.*

Basic (200) (36 of 91 SMRT Items)

Readers who have learned basic comprehension skills and strategies can locate and identify facts from simple informational paragraphs, stories, and news articles. In addition, they can combine ideas and make inferences based on short, uncomplicated passages. *Performance at this level suggests the ability to understand specific or sequentially related information.*

Intermediate (250) (4 of 91 SMRT Items)

Readers with the ability to use intermediate skills and strategies can search for, locate, and organize the information they find in relatively lengthy passages and can recognize paraphrases of what they have read. They can also make inferences and reach generalizations about main ideas and author's purpose from passages dealing with literature, science, and social studies. *Performance at this level suggests the ability to search for specific information, interrelate ideas, and make generalizations.*

Adept (300)

Readers with adept reading comprehension skills and strategies can understand complicated literary and informational passages, including material about topics they study at school. They can also analyze and integrate less familiar material and provide reactions to and explanations of the text as a whole. *Performance at this level suggests the ability to find, understand, summarize, and explain relatively complicated information.*

Advanced (350)

Readers who use advanced reading skills and strategies can extend and restructure the ideas presented in specialized and complex texts. Examples include scientific materials, literary essays, historical documents, and materials similar to those found in professional and technical working environments. They are also able to understand the links between ideas even when those links are not explicitly stated and to make appropriate generalizations even when the texts lack clear introductions or explanations. *Performance at this level suggests the ability to synthesize and learn from specialized reading materials.*

*Source: National Assessment of Educational Progress (1985, p. 15, Figure 2.3)

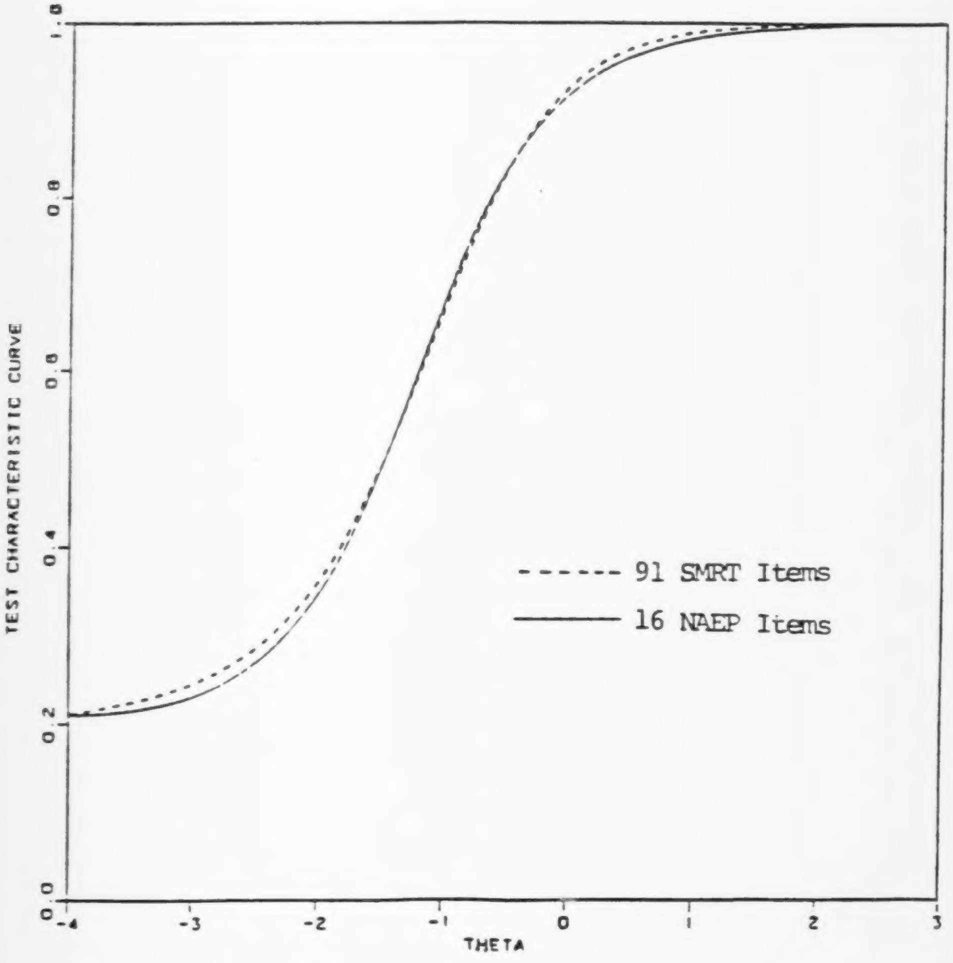


Figure 23. Plot of Test Characteristic Curves Across Theta for 91 SMRT Items and 16 NAEP Items

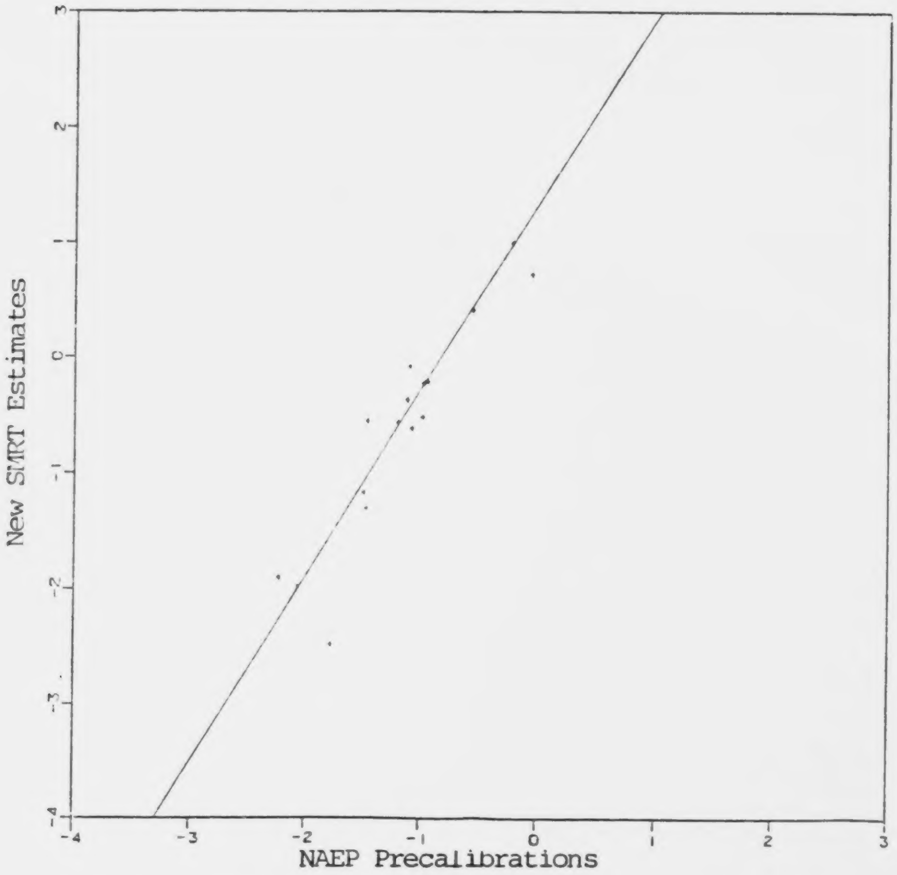


Figure 24. Relative Stability of SMRT Item Difficulty Estimates (b-values) From NAEP Precalibrations to New SMRT Estimates

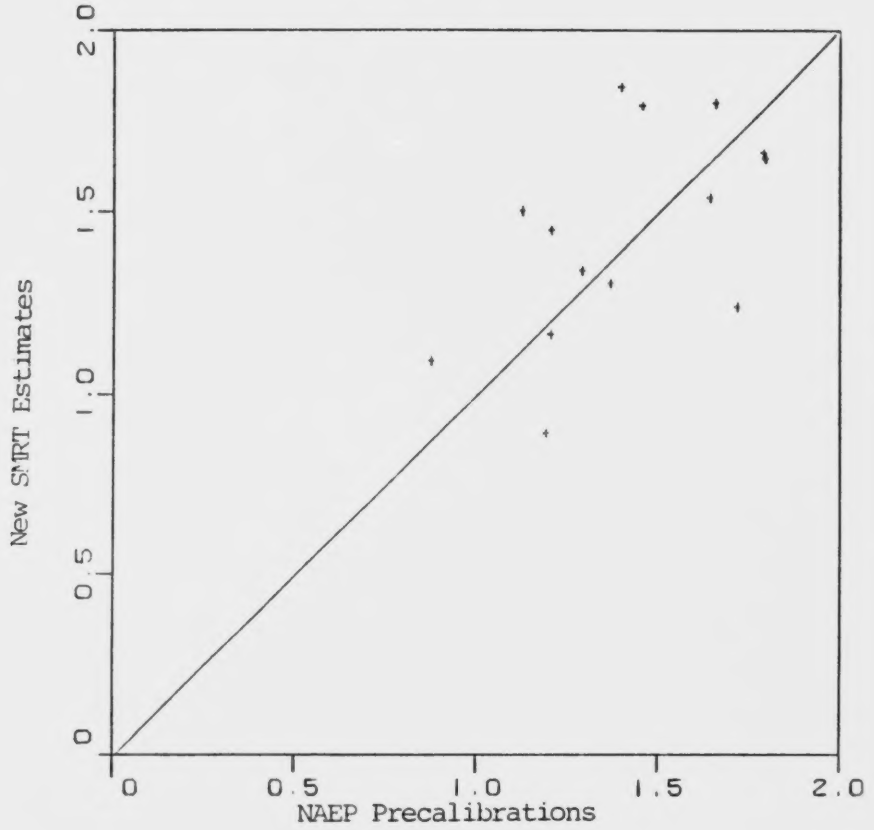


Figure 25. Relative Stability of SMRT Item Discrimination Estimates (a-values) From NAEP Precalibrations to New SMRT Estimates

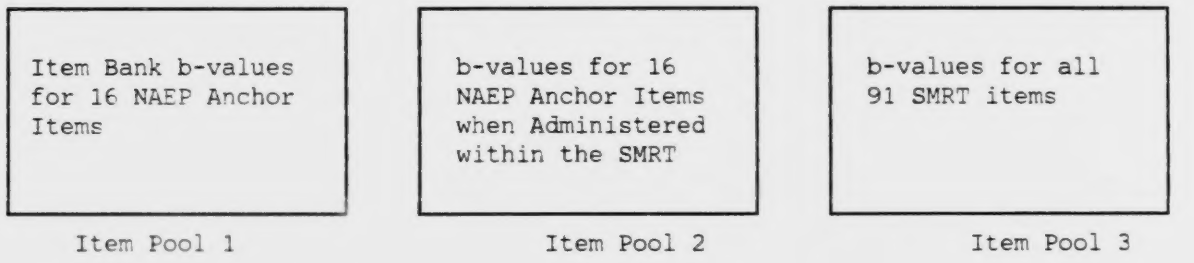

Item Bank b-values for 16 NAEP Anchor Items
b-values for 16 NAEP Anchor Items when Administered within the SMRT
b-values for all 91 SMRT items
Item Pool 1
Item Pool 2
Item Pool 3

Figure 26. Schematic of Horizontal Equating Design for SMRT in terms of NAEP.

X. ESTABLISHING MINIMUM STANDARDS FOR THE SCHOOL MASTERY OF READING TEST (SMRT)

The previous chapter discussed the relationship between the School Mastery of Reading Test (SMRT) and the National Assessment of Educational Progress (NAEP). The primary objective of that chapter was to illustrate the manner in which SMRT results could be interpreted using NAEP national norms and Levels of Proficiency. The following chapter provides Metropolitan Achievement Test (MAT) and Degrees of Reading Power (DRP) test data in addition to judgments from professional educators which might be useful in establishing (SMRT) performance standards. The current chapter does not attempt to illustrate the manner in which SMRT results could be interpreted using either MAT or DRP norms, both of which are the property of test publishers. The current chapter does, however, illustrate how MAT and DRP data (i.e., or data from other standardized reading tests) and expert judgments from educators can be used to assist in the establishment of SMRT performance standards. Such performance standards can be used to group students for appropriate instruction.

There are various procedures for establishing proficiency standards (see, for overview, Livingston & Zieky, 1982). Selection and implementation of any particular procedure should be based upon careful analysis of data, judgments, and the particular situations and potential consequences involved (Koffler, 1980). It is important to note, furthermore, that there cannot be a clear and unambiguous distinction between masters and non-masters because the underlying competency being measured (i.e., reading) is continuous and not dichotomous (Shepard, 1980).

In order to demonstrate the manner in which School Mastery of Reading Test performance criteria might be established, both empirical data and judgments of experts have been obtained. Expert judgments were provided by a Professional Panel of New York City educators involved with fourth grade students. In addition, the following data have been obtained from a total of 744 students who were administered all three standardized tests (i.e., SMRT, DRP and MAT):

1) SMRT scores of fourth graders in the nine schools tested during the second and third weeks in May 1986

2) Degrees of Reading Power (DRP) test scores of the same students tested on May 7, 1986

3) Metropolitan Achievement Test (MAT) reading scores of the same students tested on April 21 and 22, 1986

Standards For The School Mastery of Reading Test

SMRT subtest and total test results can be grouped in various ways to be useful in setting standards. Furthermore, data can be presented in tables or depicted as frequency distributions or histograms, as appropriate. The data presented in this chapter were divided into three proficiency groups based upon the New York City Board of Education's Promotional Gates criteria. However, these data could have been promulgated into a different number of groups, if educationally or psychometrically meaningful.

Table 29 summarizes both SMRT performance and professional panel judgments for groups of students whose reading performance is characterized as: below minimal competence (relatively low scoring students), minimally competent (marginal scores) and competent (relatively high scoring students). Figures 27 through 31 depict the results for the total test and each subtest. Performance standards could be established by picking points depicted by histograms or by picking points between two specific groups of histograms. For example, Figure 29 shows that Word Meaning percentages (i.e., of correct items) for the marginally competent group were approximately 75, based upon either the MAT or DRP. Therefore, 75 could be selected as the minimum standard, or some point below 75 could be selected. If a point below 75 is to be selected, it is helpful to know that student scores of the lowest achieving group averaged 59.14 and 57.19, based upon the MAT and DRP, respectively. The professional panel judgments or expected SMRT scores discussed below provide additional information potentially useful for establishing performance standards.

When making a decision about the actual subtest scores and/or total test score to be selected as standards, other factors which are of educational and psychological significance should be taken into consideration. In addition, specific standards should be promulgated as a result of a broad-based consensus provided by professional educators and parents, among others.

When establishing performance standards, it is important to know the potential citywide impact of such standards. Specifically, how many students are likely to be identified as: below minimal competence, minimally competent and competent? This is necessary in order to plan for effective use of school resources. For example, it is possible to use both the DRP and the MAT grade four spring 1986 citywide test score distributions to estimate the numbers of students in each of the three categories. These numbers are provided in Table 30.

The remainder of this chapter provides details regarding the manner in which Table 29 was promulgated. These details are provided to illustrate the research methodology utilized.

Professional Panel Expectations:

Judgments of experts were provided by a Professional Panel comprised of professional educators including teachers, assistant principals, principals, reading coordinators and curriculum supervisors (see, for discussion, Kippel and Forehand, 1986, pp. 28-50). The procedure used is based upon a rationale discussed by Angoff (1971, pp. 514-515). Panel members were asked to estimate the difficulty level of each SMRT item for each of the three hypothetical groups of students described below:

Group 1: Satisfactory or competent readers.
Students in this group:
- a) read well enough to learn from fourth grade text material in reading and other subject areas
- b) read well enough to follow instructions in workbooks, arithmetic problems, and other school work
- c) can be expected to continue to learn in the fifth grade

Group 2: Minimally or marginally competent readers.
Students in this group:
- a) have developed sufficient reading skills that they can continue to learn to read, perhaps with special help
- b) can be expected to have some difficulty with fourth grade text material, but can learn at a minimal level from such material
- c) can be expected to need continuing special help with basic reading skills in the fifth grade

Group 3: Readers below minimum competence.
Students in this group:
- a) have not achieved some or all of the basic reading skills appropriate to fourth grade
- b) cannot learn by reading fourth grade text material in reading and other subject areas
- c) cannot read sufficiently well to follow directions in workbooks and arithmetic problems

To obtain the judgments of the Professional Panel, each member was provided with a specially prepared manual which included: each SMRT item, instructions related to each SMRT item, and the following question and response categories designed to elicit their professional judgments for each item. For each of the three hypothetical groups of students, each panel member checked one of five response categories.

Professional Panel judgments or expectations of performance on each SMRT item for the three hypothetical groups of students described above were presented in the Fall 1986 Progress Report (Table 7, pages 30 through 39). Subsequently, professional panel judgments for items were combined to obtain aggregate or summary expectations for each of the four subtests and the total test. It is anticipated that standards will be based either upon item clusters or subtests, rather than upon individual items or total test scores.

The number and percent of these judgments are presented in Tables 31 through 35. For example, Table 31a presents the number of professional panel member judgments falling into each of the five columns or "expectation categories" for competent (High), minimally competent (Marginal) and below minimal competence (Low) students on the total 97 item School Mastery of Reading Test. For example, review of Table 31a indicates that the panel provided 1290 judgments or tallies indicating that 91% or more of competent readers would be expected to obtain correct scores on the 97 item test. Table 31b indicates that these 1290 tallies represent approximately 67% of the total of 1,917 judgments related to competent readers. It is apparent that the professional panel expects most competent readers to respond correctly to the total test. Review of Table 31b indicates that approximately 71% (i.e., 36% plus 35%) of panel judgments related to marginal readers were in the two columns comprising the 36% through 90% range. In effect, 36% to 90% of marginal readers would be expected to correctly answer the 97 items. Finally, 73% (i.e., 34% plus 39%) of panel judgments for below minimum competence readers were in the two columns comprising the 0% through 35% range. In other words, relatively low achieving students were expected to have difficulty correctly answering the 97 items. Subsequently, Tables 32a through 35b provide similar numbers and percentages for each of the four SMRT subtests.

Finally, in order to obtain one summary score for each of the three hypothetical groups, the percentage of judgments in each column was multiplied by a weight representing the approximate midpoint of the range at the top of that column. For example, each of the percentages in the column headed "90% or more" was multiplied by a weight of .95. Then, to obtain one summary score for each group or row, the five products were summed across the five columns. The results are presented in Table 36.

Actual Performance on the School Mastery of Reading Test

In addition to determining the expectations of professional educators, SMRT item data were obtained for competent, marginal and below minimal competence readers. These three score categories were based upon the grade four DRP Promotional Gates criterion. For example, students achieving DRP scores within one standard error either below or above the fourth grade DRP

Promotional Gates criterion were considered of minimal competence (marginal DRP scores). Students achieving DRP scores lower than one standard error below the DRP Promotional Gates criterion were considered below minimal competence (low DRP scores). Students achieving DRP scores higher than one standard error above the DRP Promotional Gates criterion were considered competent (high DRP scores).

Table 37 presents the mean and percent of correct responses for each SMRT subtest and total SMRT achieved by students in each of the three DRP - defined competence or mastery categories. The data used to derive Table 37 (i.e., the number and percent of correct responses for each SMRT item obtained by students achieving relatively low, marginal and high DRP scores) were presented in the Fall 1986 Progress Report (Table 8, pages 40-42).

In the second independent analysis summarized in Table 38, the same 744 students were again placed into one of the three categories based upon their MAT scores and the MAT Promotional Gates criterion for grade four. The SMRT subtest and total test scores for these new groups are summarized and reported in Table 38. The data used to derive Table 38 (i.e., number and percent of correct responses for each SMRT item obtained by students achieving relatively low, marginal and high MAT scores) are presented in Table 39.

Table 29

Summary of Actual and Expected School Mastery of Reading Test (SMRT) Performance for Below Minimal Competence, Minimally Competent and Competent Readers

SUBTESTS	Number of Items*	Below Minimal Competence			Minimally Competent			Competent		
		Percentage of Actual SMRT Scores Categorized By		Percentage of Professional Panel Expected SMRT Scores	Percentage of Actual SMRT Scores Categorized By		Percentage of Professional Panel Expected SMRT Scores	Percentage of Actual SMRT Scores Categorized By		Percentage of Professional Panel Expected SMRT Scores
		MAT	DRP		MAT	DRP		MAT	DRP	
WORD ATTACK	18	65.67	64.79	31.57	76.67	78.83	68.31	89.67	89.50	89.71
WORD MEANING	21	59.14	57.19	28.73	74.62	74.90	67.17	87.76	88.10	88.96
LITERAL COMPREHENSION	31	62.58	61.00	26.84	76.03	76.00	64.95	88.03	88.42	88.28
REASONING COMPREHENSION	27	54.19	54.26	22.82	66.56	65.89	60.04	79.81	80.11	84.84
TOTAL	100	61.19	60.14	26.97	73.99	74.26	64.65	86.39	86.64	87.72

* The three word recognition items are not listed as a separate subtest but are included in the total.

Table 30

Citywide Numbers of Grade Four Students in Each of Three Categories of Competence

	Estimates Based Upon DRP	Estimates Based Upon MAT
Below Minimum Competence (Relatively Low Scores)	24,710	25,427
Minimally Competent (Marginal Scores)	14,627	7,046
Competent (Relatively High Scores)	63,582	47,311
Total Number of Students in Each Distribution	102,919	79,784

Table 31

NUMBER AND PERCENTAGE OF PROFESSIONAL PANEL JUDGEMENTS OF THE PROPORTION OF STUDENTS EXPECTED TO RESPOND CORRECTLY ON THE (97 ITEM) SMRT

Table 31a: Number

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly				
	(91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	1290	572	53	2	0
Marginal	357	688	668	198	5
Low	30	167	330	642	747

* The three sample items have been eliminated from this analysis.

Table 31b: Percentage

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly				
	(91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	67%	30%	3%	0%	0%
Marginal	19%	36%	35%	10%	0%
Low	1%	9%	17%	34%	39%

* The three sample items have been eliminated from this analysis.

Table 32

NUMBER AND PERCENTAGE OF PROFESSIONAL PANEL JUDGEMENTS OF THE PROPORTION OF STUDENTS EXPECTED TO RESPOND CORRECTLY ON THE WORD ATTACK (18 ITEM) SUBTEST

Table 32a: Number

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly (91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	255	81	4	0	0
Marginal	70	146	101	23	0
Low	10	39	65	124	102

Table 32b: Percentage

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly (91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	75%	24%	1%	0%	0%
Marginal	20%	43%	30%	7%	0%
Low	4%	11%	19%	36%	30%

Table 33

NUMBER AND PERCENTAGE OF PROFESSIONAL PANEL JUDGEMENTS OF THE PROPORTION OF STUDENTS EXPECTED TO RESPOND CORRECTLY ON THE SMRT WORD MEANING (21 ITEM) SUBTEST

Table 33a: Number

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly				
	(91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	308	100	12	0	0
Marginal	97	158	121	44	0
Low	1	43	85	152	139

Table 33b: Percentage

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly				
	(91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	73%	24%	3%	0%	0%
Marginal	23%	38%	29%	10%	0%
Low	0%	11%	20%	36%	33%

Table 34

NUMBER AND PERCENTAGE OF PROFESSIONAL PANEL JUDGEMENTS OF THE PROPORTION OF STUDENTS EXPECTED TO RESPOND CORRECTLY ON THE LITERAL COMPREHENSION (31 ITEM) SUBTEST

Table 34a: Number

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly				
	(91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	428	180	11	1	0
Marginal	108	236	217	58	1
Low	8	50	111	216	235

Table 34b: Percentage

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly				
	(91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	69%	29%	2%	0%	0%
Marginal	18%	38%	35%	9%	0%
Low	1%	8%	18%	35%	38%

Table 35

NUMBER AND PERCENTAGE OF PROFESSIONAL PANEL JUDGEMENTS OF THE PROPORTION OF STUDENTS EXPECTED TO RESPOND CORRECTLY ON THE REASONING COMPREHENSION (27 ITEM) SUBTEST

Table 35a: Number

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly				
	(91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	299	211	26	1	0
Marginal	82	148	229	73	4
Low	11	35	69	150	271

Table 35b: Percentage

Hypothetical Student Performance Groups	Proportion of students expected to respond correctly				
	(91% or more)	(61-90%)	(36-60%)	(11-35%)	(10% or less)
High	56%	39%	5%	0%	0%
Marginal	15%	28%	43%	14%	0%
Low	2%	7%	13%	28%	50%

Table 36

School Mastery of Reading Test (SMRT) Panel Expectations
For Low, Marginal and High Scoring Groups

SUBTESTS	Number of Items *	Below Minimal Competence (Low)		Minimal Competence (Marginal)		Competent (High)	
		Mean and (Standard Deviation)	Percent	Mean and (Standard Deviation)	Percent	Mean and (Standard Deviation)	Percent
WORD ATTACK	17	5.37 (3.70)	31.57	11.61 (3.03)	68.31	15.25 (1.42)	89.71
WORD MEANING	21	6.03 (4.26)	28.73	14.11 (4.00)	67.18	18.68 (1.93)	88.95
LITERAL COMPREHENSION	31	8.32 (6.34)	26.83	20.13 (5.88)	64.94	27.37 (2.98)	88.28
REASONING COMPREHENSION	27	6.13 (5.06)	22.72	16.19 (5.03)	59.96	22.90 (2.80)	84.80
TOTAL	96	25.85 (19.37)	26.93	62.04 (17.94)	64.63	84.19 (9.13)	87.70

* The three word recognition items and one sample item have been eliminated from this analysis.

Table 37

School Mastery of Reading Test (SMRT) Performance For Three Degrees of Reading Power (DRP) Groups

SUBTESTS	Number of Items*	Below Minimal Competence (low DRP scores) (n=98)		Minimal Competence (marginal DRP scores) (n=123)		Competent (high DRP scores) (n=523)	
		Mean and (Standard Deviation)	Percent	Mean and (Standard Deviation)	Percent	Mean and (Standard Deviation)	Percent
WORD ATTACK	18	11.66 (2.78)	64.79	14.19 (2.82)	78.83	16.11 (2.03)	89.50
WORD MEANING	21	12.01 (3.8)	57.19	15.73 (3.09)	74.90	18.50 (2.06)	88.10
LITERAL COMPREHENSION	31	18.91 (4.79)	61.00	23.56 (4.12)	76.00	27.41 (2.72)	88.42
REASONING COMPREHENSION	27	14.65 (4.05)	54.26	17.79 (3.74)	65.89	21.63 (2.96)	80.11
TOTAL	100	60.14 (13.07)	60.14	74.26 (11.18)	74.26	86.64 (7.2)	86.64

* The three word recognition items are not listed as a separate subtest, but are included in the total.

Table 38

School Mastery of Reading Test (SMRT) Performance
For Three Metropolitan Achievement Test (MAT) Groups

SUBTESTS	Number of Items*	Below Minimal Competence (low MAT scores) (n=119)		Minimal Competence (marginal MAT scores) (n=75)		Competent (high MAT scores) (n=550)	
		Mean and (Standard Deviation)	Percent	Mean and (Standard Deviation)	Percent	Mean and (Standard Deviation)	Percent
WORD ATTACK	18	11.82 (2.99)	65.67	13.80 (2.49)	76.67	16.14 (1.98)	89.67
WORD MEANING	21	12.42 (3.96)	59.14	15.67 (2.90)	74.62	18.43 (2.07)	87.76
LITERAL COMPREHENSION	31	19.40 (4.99)	62.58	23.57 (3.41)	76.03	27.29 (2.87)	88.03
REASONING COMPREHENSION	27	14.63 (3.99)	54.19	17.97 (3.08)	66.56	21.55 (3.03)	79.81
TOTAL	100	61.19 (13.72)	61.19	73.99 (8.09)	73.99	86.39 (7.45)	86.39

* The three word recognition items are not listed as a separate subtest, but are included in the total.

Table 39

NUMBER AND PERCENT OF CORRECT RESPONSES TO EACH SCHOOL MASTERY OF READING TEST (SMRT) ITEM FOR THREE GROUPS OF STUDENTS DEFINED BY THEIR METROPOLITAN ACHIEVEMENT TEST (MAT) SCORES

(n = 744)

SMRT	Low MAT Scores (n = 119)		Marginal MAT Scores (n = 75)		High MAT Scores (n = 550)	
ITEM	Number	Percent	Number	Percent	Number	Percent
1	118	99.2	75	100.0	549	99.8
2	119	100.0	75	100.0	548	99.6
3	111	93.3	73	97.3	543	98.7
4	109	91.6	69	92.0	537	97.6
5	107	89.9	71	94.7	547	99.5
6	94	79.0	69	92.0	533	96.9
7	76	63.9	50	66.7	485	88.2
8	96	80.7	64	85.3	522	94.9
9	117	98.3	73	97.3	544	98.9
10	60	50.4	53	70.7	492	89.5
11	63	52.9	54	72.0	467	84.9
12	61	51.3	43	57.3	446	81.1
13	64	53.8	52	69.3	476	86.5
14	94	79.0	69	92.0	541	98.4
15	94	79.0	66	88.0	536	97.5
16	87	73.1	68	90.7	538	97.8
17	91	76.5	63	84.0	514	93.5
18	83	69.7	63	84.0	496	90.2
19	76	63.9	51	68.0	515	93.6
20	53	44.5	49	65.3	412	74.9
21	106	89.1	75	100.0	541	98.4
22	108	90.8	73	97.3	531	96.5
23	67	56.3	48	64.0	443	80.5
24	91	76.5	64	85.3	495	90.0
25	114	95.8	74	98.7	541	98.4
26	96	80.7	71	94.7	540	98.2
27	101	84.9	74	98.7	544	98.9
28	90	75.6	71	94.7	532	96.7
29	40	33.6	33	44.0	404	73.5
30	102	85.7	73	97.3	545	99.1
31	100	84.0	72	96.0	544	98.9
32	92	77.3	63	84.0	535	97.3
33	88	73.9	71	94.7	538	97.8
34	101	84.9	66	88.0	539	98.0
35	23	19.3	19	25.3	255	46.4
36	65	54.6	59	78.7	511	92.9
37	68	57.1	53	70.7	502	91.3

Table 39 (continued)

SMRT ITEM	Low MAT Scores (n = 119) Number	Percent	Marginal MAT Scores (n = 75) Number	Percent	High MAT Scores (n = 550) Number	Percent
38	48	40.3	42	56.0	435	79.1
39	49	41.2	50	66.7	459	83.5
40	32	26.9	38	50.7	452	82.2
41	66	55.5	51	68.0	469	85.3
42	94	79.0	70	93.3	489	88.9
43	76	63.9	62	82.7	487	88.5
44	67	56.3	57	76.0	470	85.5
45	85	71.4	69	92.0	491	89.3
46	85	71.4	63	84.0	501	91.1
47	89	74.8	67	89.3	525	95.5
48	57	47.9	46	61.3	455	82.7
49	75	63.0	49	65.3	451	82.0
50	77	64.7	63	84.0	511	92.9
51	78	65.5	59	78.7	508	92.4
52	92	77.3	64	85.3	534	97.1
53	81	68.1	58	77.3	513	93.3
54	52	43.7	42	56.0	453	82.4
55	78	65.5	65	86.7	524	95.3
56	73	61.3	59	78.7	522	94.9
57	25	21.0	34	45.3	336	61.1
58	63	52.9	51	68.0	443	80.5
59	81	68.1	57	76.0	468	85.1
60	77	64.7	61	81.3	504	91.6
61	74	62.2	64	85.3	494	89.8
62	81	68.1	66	88.0	527	95.8
63	77	64.7	57	76.0	456	82.9
64	34	28.6	39	52.0	331	60.2
65	50	42.0	40	53.3	427	77.6
66	48	40.3	40	53.3	435	79.1
67	39	32.8	41	54.7	436	79.3
68	53	44.5	54	72.0	478	86.9
69	84	70.6	65	86.7	517	94.0
70	94	79.0	62	82.7	534	97.1
71	82	68.9	58	77.3	506	92.0
72	43	36.1	47	62.7	462	84.0
73	72	60.5	61	81.3	523	95.1
74	97	81.5	72	96.0	530	96.4
75	79	66.4	63	84.0	480	87.3
76	85	71.4	58	77.3	476	86.5
77	73	61.3	66	88.0	516	93.8
78	80	67.2	62	82.7	478	86.9
79	64	53.8	55	73.3	472	85.8
80	82	68.9	61	81.3	505	91.8
81	44	37.0	28	37.3	401	72.9
82	28	23.5	30	40.0	425	77.3
83	31	26.1	35	46.7	424	77.1
84	29	24.4	24	32.0	328	59.6
85	47	39.5	45	60.0	454	82.5

Table 39 (continued)

SMRT	Low MAT Scores (n = 119)		Marginal MAT Scores (n = 75)		High MAT Scores (n = 550)	
ITEM	Number	Percent	Number	Percent	Number	Percent
86	28	23.5	18	24.0	280	50.9
87	48	40.3	43	57.3	464	84.4
88	93	78.2	67	89.3	537	97.6
89	94	79.0	70	93.3	530	96.4
90	92	77.3	66	88.0	535	97.3
91	89	74.8	56	74.7	459	83.5
92	84	70.6	63	84.0	507	92.2
93	24	20.2	18	24.0	301	54.7
94	94	79.0	66	88.0	519	94.4
95	23	19.3	16	21.3	324	58.9
96	88	73.9	67	89.3	475	86.4
97	50	42.0	42	56.0	403	73.3
98	26	21.8	27	36.0	264	48.0
99	31	26.1	22	29.3	205	37.3
100	23	19.3	29	38.7	313	56.9

Figure 27
Percent of Total School Mastery of Reading Test Items Correct for Low, Marginal and High Score Groups

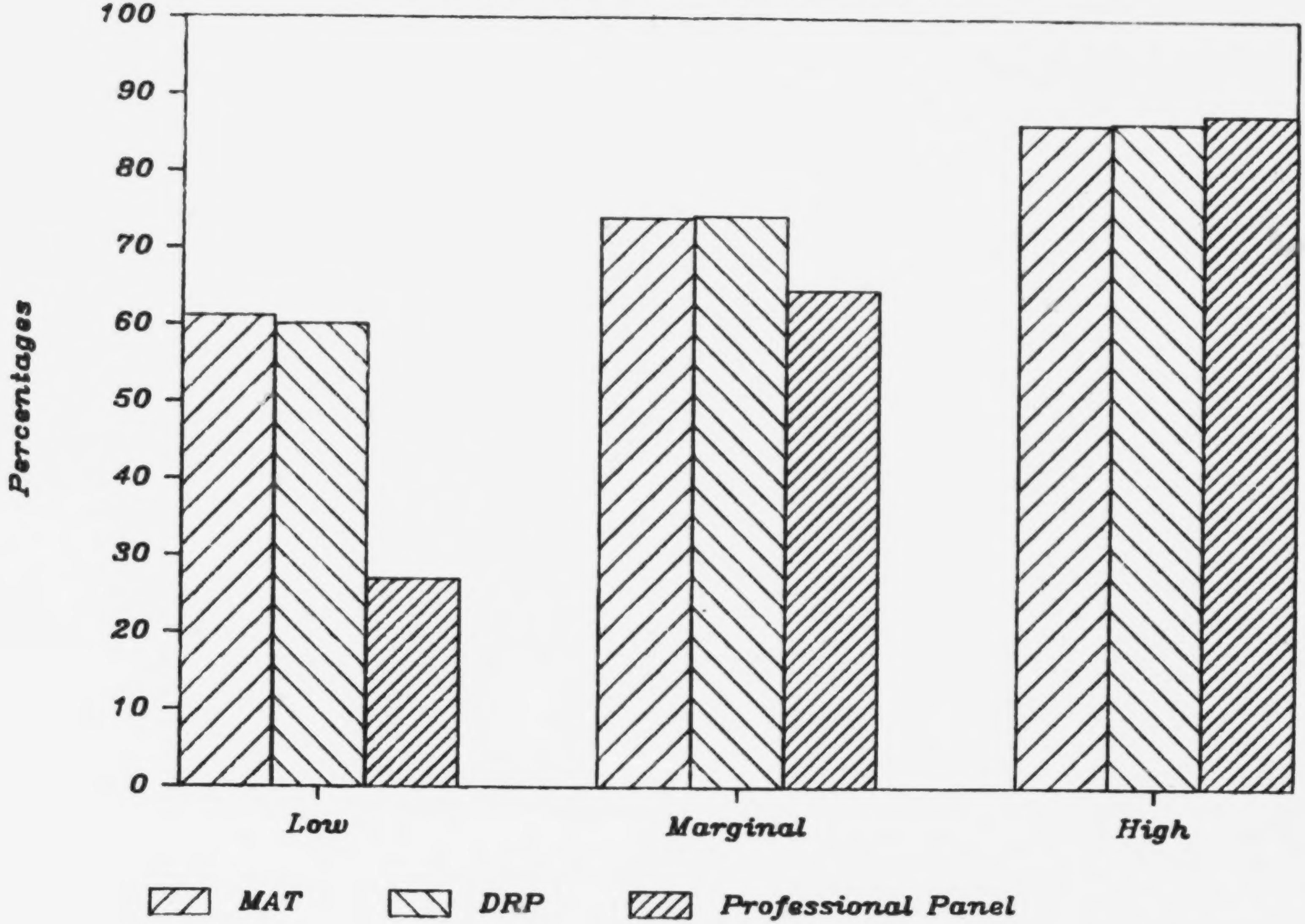


Figure 28

Percent of Word Attack Subtest Items Correct for Low, Marginal and High Score Groups

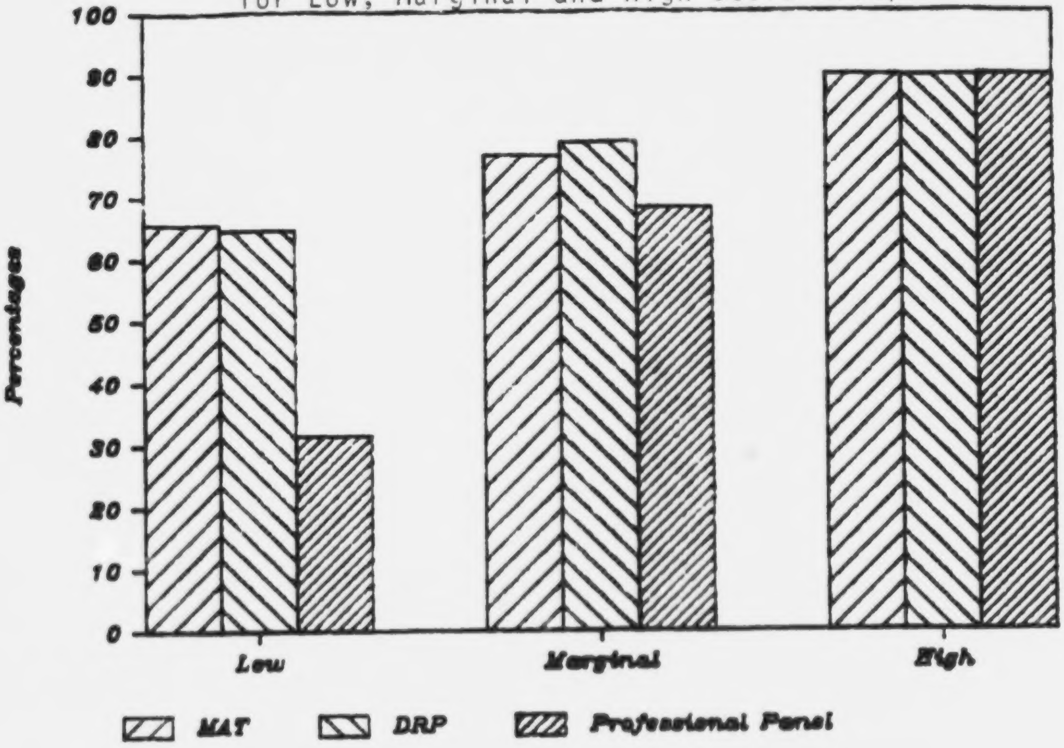


Figure 29

Percent of Word Meaning Subtest Items Correct for Low, Marginal and High Score Groups

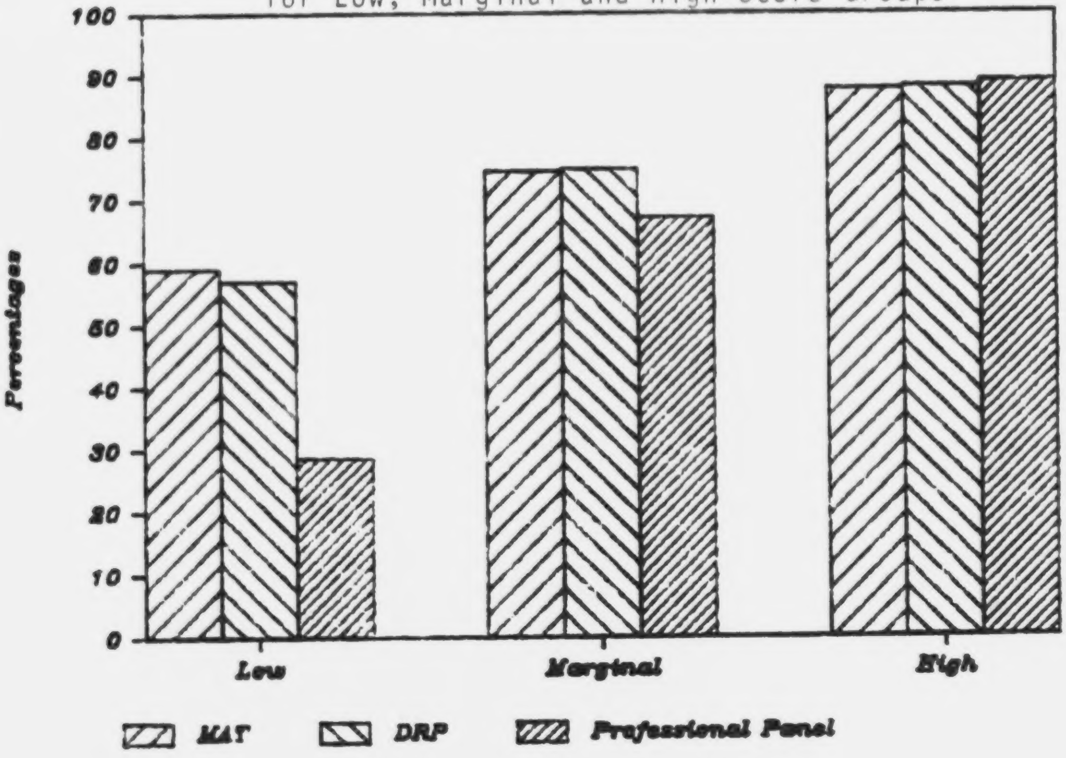


Figure 30

Percent of Literal Comprehension Subtest Items Correct for Low, Marginal and High Score Groups

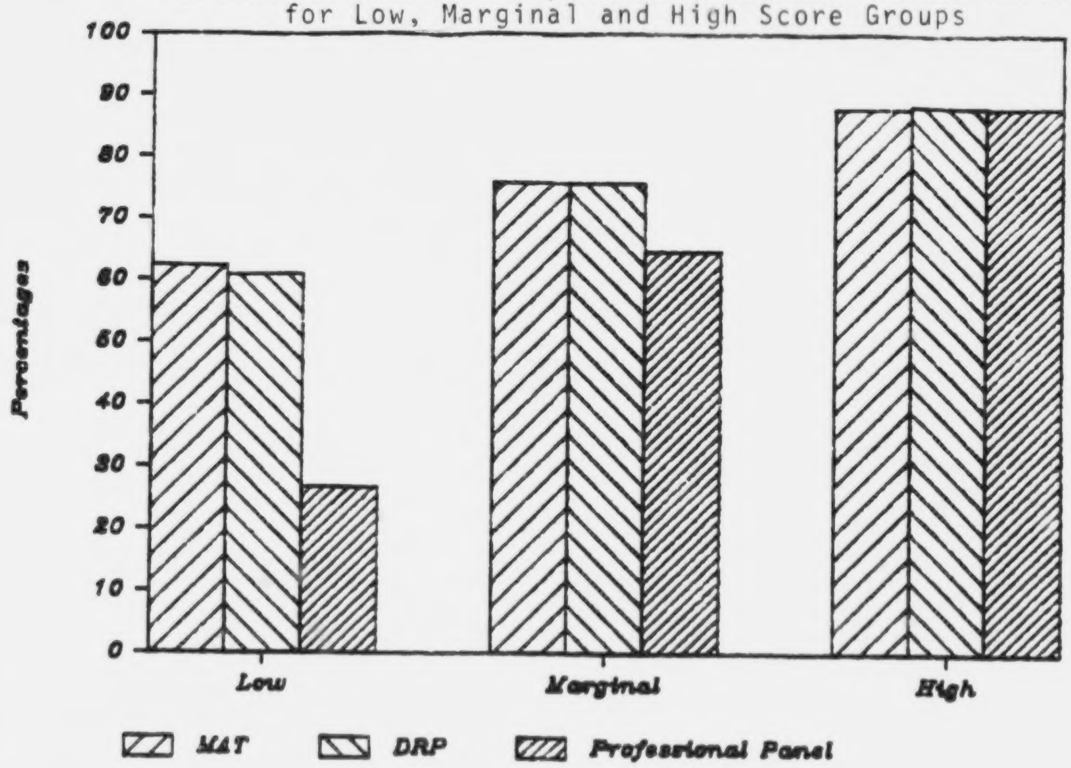
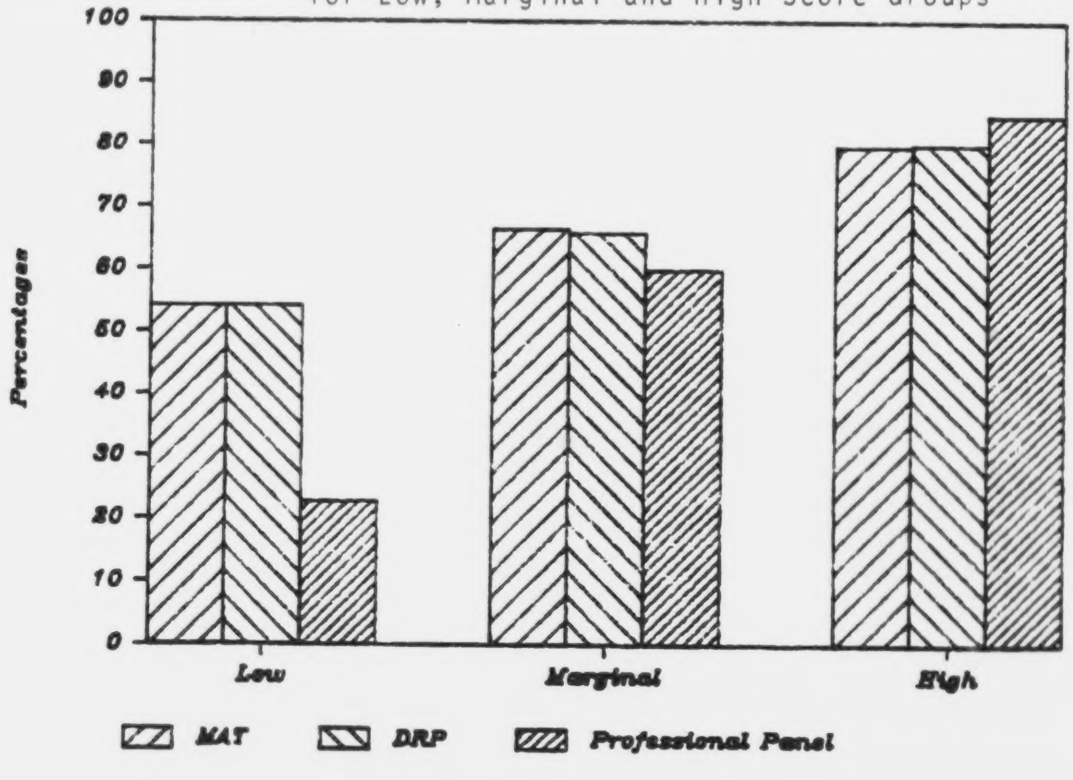


Figure 31

Percent of Reasoning Comprehension Subtest Items Correct for Low, Marginal and High Score Groups



XI. RELATIONSHIP BETWEEN SMRT-STEPS AND CIMS-CA PROJECT

The table on the next page presents an overview of the relationship between SMRT-STEPS and CIMS-CA. It is apparent that the two projects are very different in nature and scope.

The goals of the New York City Board of Education's Comprehensive Instructional Management System-Communication Arts (CIMS-CA) project are to develop a holistic communication arts curriculum for kindergarten through eighth grade, a corresponding test component, and a computer management system. The curriculum and test components, developed by teachers, integrate the four content areas in communication arts -- reading, writing, listening, and speaking. For the 1985-86 school year, a drama component was added to the curriculum. The CIMS-CA project is being implemented in Community School Districts 8, 9, 11, 15, 17 and 30. The objectives of the SMRT-STEPS Project have been delineated earlier in the chapter entitled: "Brief Description of the School Mastery of Reading Test System to Enhance Progress of Schools (SMRT-STEPS) Project."

Table 40

The Relationship Between SMRT-STEPS and CIMS-CA

	SMRT-STEPS	CIMS-CA
Curriculum Foundation	Standard Citywide Curriculum (including Minimum Teaching Essentials). Additional Experimental Edition Developed Using Basal Reader	Most Assessment Specific to CIMS-CA Objectives (which are based upon Minimum Teaching Essentials and New York State syllabus)
Scope and Cost	Assessment Component Only	Provides Curriculum Component and Teacher Workshops in Addition to Assessment Component.
Number of Workshops Required	Minimum Number of Professional Teacher and Supervisor Workshops Required to Review Curriculum Relevance and all Procedures and Products	Ongoing Teacher Training Workshops Required for Test Development and to Discuss Procedures for Administration and Scoring
Norm-Referenced Interpretation	Linked to National Assessment of Educational Progress National Norms	Not Provided
Scaled Scores	Based Upon Item Response Theory Calibration	Not Provided
Mastery Scores and/or Performance Standards	Based Upon Teacher and Supervisor Judgments and Relationships With Degrees of Reading Power (DRP) Metropolitan Achievement Test (MAT) and/or National Assessment of Educational Progress (NAEP)	Based Upon Teacher and Supervisor Judgments
Test Item Development	Items Developed By Reading Specialists or Obtained From Existing Item Banks and Reviewed By New York City Teachers and Supervisors	Items Developed by Teachers and Supervisors Working With Reading Specialists

Table 40 (continued)

	SMRT-STEPS	CIMS-CA
Number of Test Items	Plan to Develop Extensive Item Banking as Required for Ongoing Generation of Alternate Forms.	Limited Number of Items for Each Objective and Theme. Potential for Expanded Bank of Items.
Relationships With other Standardized Tests (e.g., Degrees of Reading Power, Metropolitan Achievement Test, National Assessment of Educational Progress)	Regression Analyses Conducted. Citywide Projections Being Estimated. Item Response Theory Analyses Conducted for SMRT and NAEP.	Not Provided
Bias Review	During Item Development. Subsequent review by Professional Panel of New York City Teachers and Supervisors	During Item Development. Subsequent Review by University Consultants. Revision by Teachers and Supervisors on Advice of University Consultant.
Scoring Method	Hand- or Machine-Scored	Hand- or Machine-Scored
Timed/Untimed (Speed vs. Power Test)	Untimed - Estimated Time Guidelines Provided	Untimed
Institutional Relationship	Consortium with Educational Testing Service of Princeton, New Jersey.	Reviewed by Consultant from New York State Education Department
Reports of Results	Reports Include Individual Student Listings, Class and Grade Reports.	Reports Include Individual Student and Class Reports for Reading, Listening-Speaking and Writing. Additional Archive Reports Show Student Test Results Through the Grades.
Test Security	Secure Test	Non-Secure Test

XII. REVIEW OF OTHER STANDARDIZED READING TESTS

A review of standardized reading tests frequently administered in New York City schools was conducted to determine if such instruments might be appropriate, cost-effective and useful for improving New York City schools. Both oral and written standardized reading tests were reviewed to determine if any might be useful, in particular, as the SMRT-STEPS assessment component. Among those tests reviewed, no currently existing standardized reading test was found to be an adequate substitute for a new test based specifically upon New York City curriculum.

A test with most of the following characteristics was sought:

1. Valid for group administration
2. Criterion-referenced with norm-referenced interpretation
3. Appropriate for New York City Communication Arts - Reading curriculum
4. Both content and concurrent validity demonstrated
5. Reliability demonstrated
6. Free of test bias
7. Machine-scorable answer sheets
8. Untimed test administration
9. Item bank or additional items available for customization
10. Teachers involved in test development
11. Mastery criteria established
12. Prescriptive instructional strategies available

Tests were selected for review based upon the recommendations of New York City Board of Education curriculum, instruction and testing specialists. In addition, Buros' "Mental Measurements Yearbook," and books on assessment and professional journals were consulted both to identify tests for consideration and as a source of critical reviews. Also, the "Test Resource Book" was carefully examined. This publication was prepared by the New York City Board of Education's Division of Special Education and presents reviews of standardized tests used in New York City schools.

Project staff reviewed standardized reading tests which are frequently administered in New York City public schools using both "Instructions for completing a test review" (see Appendix B) prepared by the Division of Special Education and "..supplemental guidelines.." (see Appendix C) prepared by project staff. The resulting overview of frequently used tests was prepared and is presented in Table 41. More detailed reviews were presented to the Division of Special Education for inclusion in subsequent editions of the "Test Resource Book."

It is noted that tests administered as part of the annual spring citywide reading testing program were not included in these reviews. The citywide reading testing program is administered primarily to obtain norm-referenced information to meet legal requirements to rank schools for teacher selection purposes. In contrast, the primary purpose of this search was to attempt to locate an instructionally useful criterion-referenced test which is strongly related to New York City curriculum and which may serve as an adjunct to any New York citywide reading test.

Overview of Frequently Administered Tests

The first row of Table 41 specifies whether the test is individually or group administered. Group tests are more cost-effective and practical than individual tests. When testing for program evaluation, screening, and/or program planning, the expense and loss of instructional time required for individually administered tests may not be justified in terms of the information desired. Individually administered tests may, in some instances, provide more valid results. Table 41 indicates that four of the seven tests are individually administered tests only. The TORC may be either individually or group administered and the PRI/RS and SDRT are group administered only.

The second row in the table reports on the nature of the test materials. In general, machine-scorable answer sheets with reusable test booklets are more desirable than consumable test booklets because they are cost-effective. In two of the seven instances, the tests require responses directly in the test booklet in such a way that the booklets are consumed and cannot be used a second time. In the other five instances, answer sheets with reusable test booklets for specific test levels are provided for at least some levels.

The third row reports on the scoring method. During the scoring process, the number of items which are correct or incorrect is obtained and: 1) publisher-developed tables are used to translate raw scores (i.e., number of items correct) into standard scores, percentiles, age or grade equivalents; and/or 2) mastery levels are determined; and/or 3) profile charts are established by following the scoring procedures outlined in the test manual.

Hand-scoring provides almost immediate results which may result in maximum instructional usefulness. However, hand-scoring is only as accurate as the scoring skills of the examiner and may be complicated, tedious, time-consuming and somewhat more subject to error than machine-scoring. In five instances reported in Table 41, tests can be hand-scored only. Machine-scanning and -scoring is particulary desirable for large-scale testing because it is relatively accurate and cost-effective. The PRI/RS and SDRT may be either hand- or machine-scored.

The fourth row reports on the availability of supplementary items to enable the customization or tailoring of the publisher's shelf test to meet specific school needs. Item banks catalogued by test objective, for example, provide this potential. None of the tests provide supplemental items, and/or objectives which enable "customization" for local use.

The fifth row reports on the time allotted to administer specific tests, excluding the time needed for scoring and interpretation. Five tests are untimed in that students are permitted to work at their own pace. This is a desirable feature and allows for individual student response rate differences. Consequently, student frustration may be reduced and subsequent test performance may be a more valid measure of student achievement. The Gates-McKillop and SDRT contain subtests which state exact time limits. For example, the "Words: Flash" subtest in the Gates-McKillop is a timed word identification test that requires the use of a tachistoscope.

The sixth and seventh rows report on whether the test publishers claim to provide norm-referenced (NRT) or criterion-referenced (CRT) test interpretations. It is noted that tests are developed as norm-referenced or criterion referenced, but not both. However, a criterion-referenced test may to some extent, provide a norm-referenced interpretation. Similarly a norm-referenced test may, to some extent, provide a criterion-referenced interpretation.

A norm-referenced interpretation provides a means of comparing a student's performance to that of other students. Results may be in terms of standard scores, normal curve equivalents, percentiles, and age or grade equivalents. A criterion-referenced interpretation addresses the assessment of particular skills in terms of levels of mastery. Results usually indicate mastery, partial mastery or non mastery of specific skills. Such scores are particularly useful for instructional planning, screening and program evaluation.

Four tests provide only norm-referenced interpretation. The SDRT and Woodcock are categorized as providing both norm- and criterion-referenced interpretations. In addition to providing criterion-referenced information, the PRI/RS offers norm-referenced interpretations based upon correlations with the California Achievement Tests (CAT C & D) and the Comprehensive Tests of Basic Skills (CTBS U and V).

The eighth row reports on how mastery levels were determined. In general, norm-referenced tests do not report mastery levels. In these instances, results are reported in terms of raw scores, stanines, scaled scores, percentiles, age/grade equivalents, normal curve equivalents, and/or quotients. For three of the tests reviewed, mastery scores are provided. The PRI/RS reports results as raw scores, and indicates mastery, partial mastery, or non mastery of specific skills. Test results of the SDRT and Woodcock are reported as raw scores, grade scores, percentile ranks,... as well as relative mastery levels or "Progress Indicator" scores.

The ninth row reports on the type of validity addressed - Concurrent, Construct, Content, Predictive. Concurrent Validity "refers to how accurately a student's current test score can be used to estimate the current criterion score" (Salvia & Ysseldyke, p.135). It is usually demonstrated by comparing test results with test scores of similar tests that are presumed to be valid. Concurrent validity was reported in six of the seven tests. Construct Validity is concerned with the meaning of the test. A construct is a psychological term referring to something that is not directly observable, but is literally constructed by a person to account for regularities or relationships observed. The construct validity of a test reflects the positive evidence collected that the test is in fact assessing the hypothesized construct. Content Validity refers to the degree to which we can generalize from the sample of items in a test, to a specified domain or universe of items. It reflects how well a test represents that which expert judgment would consider to be important knowledge or skill. Predictive Validity reflects how well a particular test or set of items predicts the criterion. It tells us the degree to which we can predict future performance on the basis of current test scores.

The tenth row specifies whether it is reported that the test was developed with teacher input. In general the involvement of teachers in the test development process may increase the meaningfulness and instructional usefulness of the test results. Six tests did not specify teacher participation in test development. The seventh test, the PRI/RS, indicated that the test was an outgrowth of research on popular basal reading programs. This research was conducted by developmental and diagnostic reading specialists. In addition, pre- and post-test questionnaires were completed by teachers involved in validation studies.

The eleventh row reports on how the individual tests relate to the New York City Board of Education's Communication Arts - Reading curriculum as outlined in the "Minimum Teaching Essentials" (MTE). None of the test manuals include specific references to the MTE. Only the authors of the PRI/RS and SDRT state that attempts were made to make objectives consistent with common reading curricula, but no specific school districts were mentioned. The sequence of reading skills that would be listed in the MTE's is not applicable to the Gilmore and Gray which are oral reading tests.

Problems occur when reading tests and curricula are not congruent. Students obtaining instruction within a specific curriculum learn specific skills relevant to that particular curriculum. Students obtaining instruction based upon different curricula may perform differently on the same standardized test. Obviously, a test should measure what has been taught. If there is a difference between what has been taught and what is tested, that test is not a valid measure of instruction. Consequently, it is essential to use standardized tests based upon New York City curriculum to assess the progress of children who were provided instruction based upon that New York City curriculum.

The twelfth row reports on the existence of procedures implemented to minimize test bias. For example, the manual of the PRI/RS specifies that test items are free of any cultural, racial, gender, SES, regional, age, and handicapping condition bias due to the implementation of procedures designed to minimize such bias. In contrast, the other test manuals did not specifically address this issue.

The thirteenth row reports on prescriptive instruction strategies. It is desirable for a test manual to provide specific instructional strategies since it increases the usefulness of test results and the test as a whole. Five tests do not provide such strategies. However, the SDRT includes the "Handbook for Instructional Techniques and Materials" and the "Manual for Interpreting." The PRI/RS incorporates a variety of supplemental materials such as the "Teacher Resource File" with lesson plans to teach specific skills and the "Tutor Activities" student worksheets.

In summary, each test showed strengths in some of the characteristics sought. However, no one test fulfilled the major characteristics required to be an adequate substitute for a new test based specifically upon New York City curriculum.

Table 41

Overview of Frequently Used Standardized Reading Tests

CRITERIA	Gates-McKillop-Horowitz Reading Diagnostic Test 1981 Edition	Gilmore Oral Reading Test 1968 Edition	Gray Oral Reading Test Revised (GORT-R) 1968 Edition	PRI Reading System (PRI/RS) 1980 Edition	Stanford Diagnostic Reading Test (SDRT) 1984 Edition	Test of Reading Comprehension (TORC) 1986 Edition	Woodcock Reading Mastery Tests 1973 Edition
1-Administration Method	individual	individual	individual	group	group	individual or group	individual
2-Nature Of Test Materials	test booklets	students answer orally and teacher records answer in answer sheet	students answer orally and teacher records answer in test booklet	test booklets or answer sheets (depending on grade level)	test booklet or answer sheets (depending on grade level)	test booklets or answer sheets	test booklets or answer sheets
3-Scoring Method	hand-scored	hand-scored	hand-scored	hand- or machine scored	hand- or machined scored	hand-scored	hand-scored
4-Availability Of Items For Customization	not available	not available	not available	not available	not available	not available	not available
5-Timed/ Untimed	depends on subtest	untimed	untimed	untimed	depends on subtests	untimed	untimed
6-Norm-Referenced Interpretation	yes	yes	yes	yes (based upon other test norms)	yes	yes	yes
7-Criterion-Referenced Interpretation	no	no	no	yes	yes	no	yes
8-Mastery Determination	not appropriate	not appropriate	not appropriate	mastery scores provided	mastery scores provided	not appropriate	mastery scores provided
9-Validity Addressed	Concurrent	Concurrent (with earlier edition)	Concurrent Content Construct	Concurrent Content	Concurrent Content	Concurrent Content Construct	Content Construct Predictive
10-Teacher Input In Test Design And Development	not specified	not specified	not specified	reading specialists only	not specified	not specified	During item Development Process
11-Relation To Minimum Teaching Essentials	not specified	not applicable	not applicable	not specified	not specified	not specified	not specified
12-Efforts To Eliminate Test Bias	not specified	not specified	not specified	were reported	Bias Panel	not specified	Bias Panel
13-Provides Prescriptive Instructional Strategies	not provided	not provided	not provided	provided	provided	not provided	not provided

BEST COPY AVAILABLE

XIII. SUMMARY OF FINDINGS AND ACCOMPLISHMENTS

In both fall 1986 and spring 1987, the School Mastery of Reading Test (SMRT) was administered to both third and fourth graders in nine Comprehensive Assessment Report (CAR) elementary schools in three Community School Districts (see, for discussion, Chapters III through VI). Both cross-sectional and longitudinal data were analyzed.

The following results suggest the validity of SMRT (see, for additional discussion of validity, Kippel and Forehand, 1986, pp. 51-56):

- In both grades three and four, scores from the spring 1987 SMRT administration were consistently higher than scores from the fall 1986 test administration. In addition, grade four test scores were generally higher than those for grade three (see, for discussion of results, Chapter VI)
- Both third and fourth grade students obtained the highest percentage of items correct on the word attack subtest and lowest on the reasoning comprehension subtest. This is consistent with curriculum and instruction emphasis
- In both grades three and four, test score distributions especially in spring were negatively skewed indicating a "piling up of scores" at the high end of the score distribution. This is the type of test score distribution expected from a mastery test related to curriculum and is administered at the end of the academic year
- Correlational evidence supports the validity of the SMRT subtests (see Chapter VIII)

In addition the following are noted :

- Reliability estimates for grades three and four, for both fall and spring, provide support for the contention that SMRT can be used reliably (see Chapter VII)
- A prototype of SMRT New York City norms has been established by generating percentile and stanine norms using SMRT raw scores (see, for discussion, Kippel and Forehand, 1986, pp. 21-25)
- The validity of calibrating SMRT items onto the National Assessment of Educational Progress (NAEP) scale has been demonstrated. Consequently, SMRT results can be interpreted

with respect to NAEP national norms and performance standards. Furthermore, SMRT items can be replaced with comparable NAEP items (see Chapter IX)

- A framework for establishing SMRT performance standards or levels of proficiency is illustrated using the Metropolitan Achievement Test (MAT), Degrees of Reading Power (DRP) test data and expert judgments from a professional panel of New York City educators (see Chapter X)

- A test administration manual was developed and used successfully by third and fourth grade teachers, with no advanced test administration training

- A comprehensive review of standardized reading tests frequently administered in New York City schools revealed that no currently existing test was an adequate substitute for a new test based specifically upon New York City curriculum (see Chapter XII)

- Survey results provided by field practitioners, including both teachers who administered SMRT and Professional Panel members, reflected very favorably on the potential usefulness of SMRT (see, Kippel and Forehand, 1986, pp. 48-49 and 54-55; also see, Kippel and Forehand, 1987, pp. 36-38)

- It has been demonstrated that SMRT can be administered cost-effectively by developing re-usable test booklets and using machine-scannable answer sheets (see, for discussion of machine-scoring, Kippel and Forehand, 1986, pp. 14-15; also see, for discussion of answer key, Kippel and Forehand, 1987, pp. 39-40)

- Assessment of the relationship between SMRT-STEPS and the Comprehensive Instructional Management System-Communication Arts (CIMS-CA) projects revealed that SMRT-STEPS and CIMS-CA are very different in nature and scope (see Chapter XI)

- Project staff maintained ongoing liaison with School Improvement Program (SIP) staff regarding the relationship between SMRT-STEPS and current New York City school improvement efforts

- A professional panel comprised of New York City Educators was convened to provide a broader perspective to the project and to increase the usefulness of all aspects of SMRT-STEPS. This panel reviewed SMRT for potential bias and provided judgments related to mastery criteria. In addition, panel member opinions were obtained regarding the usefulness of types of test scores and standardized tests. (see, for discussion, Kippel and Forehand, 1986, pp. 26-50)

- A funding proposal is being developed in collaboration with the Educational Testing Service (ETS) in Princeton, New Jersey, for potential submission to federal, state government and private foundations. This will include provision for a sophisticated computerized item-banking system to facilitate test development, item storage and record-keeping

BIBLIOGRAPHY

Angoff, W.H. (1971). Scales, norms and equivalent scores. In R.L. Thorndike (Ed.). Educational measurement (2nd ed.). Washington, DC: American Council on Education, pp. 508-600.

Brussis, Anne M., & Chittenden, Edward A. (1987). Research currents: What the reading tests neglect. Language Arts, 64, 302-308.

Flood, James & Lapp, Diane. (1986). Types of text: The match between what students read in basals and what they encounter in texts. Reading Research Quarterly, 21, 284-297.

Goodman, Yetta M. (1985). Kidwatching: Observing children in the classroom. In Angela Jaggar & M. Trika Smith-Burke, Observing the language learner. Newark, DE: International Reading Association, 9-18.

Guilford, J.P. (1954). Psychometric methods (2nd ed.). New York: McGraw-Hill.

Johnston, P. (1983). A cognitive basis for the assessment of reading comprehension. Newark, DE: International Reading Association.

Kerlinger, F.N. (1973). Foundations of behavioral research (2nd ed.). New York: Holt, Rinehart & Winston.

Kippel, G.M., & Forehand, G.A. (1986). SMRT-STEPS; School Mastery of Reading Test System To Enhance Progress of Schools; Fall, 1986 progress report. Brooklyn, New York: Board of Education of New York City, Office of Educational Assessment.

Kippel, G.M., & Forehand, G.A. (1987). SMRT-STEPS; School Mastery of Reading Test System To Enhance Progress of Schools; Fall 1986 field test results. Brooklyn, New York: New York City Board of Education.

Koffler, S.L. (1980). A comparison of approaches for setting proficiency standards. Journal of Educational Measurement, 17, 167-178.

Komoski, P.K. (1987). Proposal to thirty-five corporate foundations to date. Water Mill, New York: Educational Products Information Exchange.

Livingston, S.A., & Zieky, M.J. (1982). Passing scores: A manual for setting standards of performance on educational and occupational tests. Princeton, New Jersey: Educational Testing Service.

BIBLIOGRAPHY (continued)

McClung, M.S. (1978). Are competency testing programs fair? legal?: Phi Delta Kappan, pp. 397-400.

Messick, S. (1985). Progress toward standards as standards for progress: A potential role for NAEP. Educational Measurement: Issues and Practice, 4, 16-19.

Messick, S., Beaton, A., & Lord, F. (1983). National Assessment of Educational Progress reconsidered: A new design for a new era. (NAEP Report 93-1). Princeton, New Jersey: National Assessment of Educational Progress at Educational Testing Service.

Mitchell, J.V., Jr. (Ed.). 1985. The ninth mental measurements yearbook. Lincoln, Nebraska: The University of Nebraska Press.

National Assessment of Educational Progress (1985). The reading report card: Progress toward excellence in our schools; Trends in reading over four National Assessments, 1971-1984 (Report No. 15-R-01). Princeton, New Jersey: National Assessment of Educational Progress at Educational Testing Service.

New York City Board of Education (1968). Sequential levels of reading skills; Prekindergarten-grade 12. Brooklyn, New York: Division of Curriculum and Instruction.

New York City Board of Education (1969). Handbook for language arts; Grades three and four. Brooklyn, New York: Division of Curriculum and Instruction.

New York City Board of Education (1980). Minimum teaching essentials; Grades 3-5. Brooklyn, New York: Division of Curriculum and Instruction.

New York State Education Department (1986). English language arts syllabus, K-12; Field test edition/1986. Albany, New York: The University of the State of New York, Bureau of English and Reading Education, Bureau of Curriculum Development.

Nunnally, J.C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Omanson, R. (1982). The relation between centrality and story category variation. Journal of Verbal Learning and Verbal Behavior, 21, 326-337.

Pearson, P. David (Ed.). (1984). Handbook of reading research. New York: Longman.

BIBLIOGRAPHY (continued)

Rusch, G.M., & Stoddard, G.D. (1927). Tests and measurements in high school instruction. Yonkers, NY: World Book.

Salvia, J., & Ysseldyke, J.E. (1985). Assessment in special and remedial education (3rd ed.). Boston: Houghton Mifflin.

Schank, R.C. (1975). The structure of episodes in memory. In D.G. Bobrow and A. Collins (Eds.). Representation and understanding: Studies in cognitive science. New York: Academic Press.

Shepard, L. (1980). Technical issues in minimum competency testing. In D.C. Berliner (Ed.). Review of research in education (Volume 8). Itasca, Il: F.E. Peacock.

Thorndike, R.L., & Hagen, E.P. (1977). Measurement and evaluation in psychology and education (4th ed.). New York: John Wiley & Sons.

Wilson, S.M., & Hiscox, M.D. (1984). Using standardized tests for assessing local learning objectives. Educational Measurement: Issues and Practices. pp. 19-22.

Wixson, Karen K., & Peters, Charles W. (1987). Comprehension assessment: Implementing an interactive view of reading. Educational Psychologist (in press).

Appendix A

ITEMS OBTAINED FROM THE NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (NAEP)

Question 25: A dog lying on top of doghouse.

Question 33: Puzzle about chair.

Questions 34-36: International News: Naomi James. Reprinted by permission of Radosevich, Mayer and Associates.

Questions 37-40: What is Quicksand? From World and Space (1976). Volume 4 of Childcraft. The How and Why Library. Field Enterprises Educational Corporation.

Questions 81 & 82: Passage about a dog and his shadow. The Dog and the Shadow. From Aesop's Fables, Harper and Row Publishers, Inc. (1927).

Questions 83-85: Reading about "crickets." Special permission granted by Would You Believe Published by Xerox Education Publications (1974) Xerox Corp.

Questions 86 & 87: Reading passage about the origin of the sandwich. Special permission granted by Would You Believe Published by Xerox Education Publications (1974) Xerox Corp.

INSTRUCTIONS FOR COMPLETING A TEST REVIEW[1]

The following indicates the kind of information which should be included in each section and some standard statements that may be applicable.

Title ____________________

Author ____________________

Copyright Date ____________

Age/Grade ________________

Time ____________________

Type ____________________

Intended Purpose of Test

(Stated purpose as included in the manual; specific student population for whom test was designed including age/grade range)

Suggested Statements

The _________ is designed to measure _________________.

The ___________ is designed to assess ________________.

The _________ is designed to gather information on ______.

The _______ is designed to be used with (young) students to assess _____________________.

Description

(Details of test content should be listed- for example objectives or subtest titles and number and types of tasks; criterion for mastery; number of test forms; nature of the materials--flip over kit)

Objectives Types of Tasks (Number of items)

A score of ______ is required for mastery on each subtest.

The ______ consists of ______.

It includes __________________.

The ______ is a kit consisting of __________________________.

The ______ has ________ items which ________________________.

The _______ has _____ objectives grouped into _________ subtests of _______________ items.

The _________ is divided into __________ areas of ________.

The _____________ is packaged as a _______________________.

Test Administration

(Rules for administration; training; examiner qualifications/training/experience; individualized/group; special supplies-time clocks, paper, pencils.)

The __________ is administered (individually/group).

The examiner required (special/no) training.

It takes (time) to administer.

The administrator requires _______________.

In order to give _____________, the following materials should be available: ________________.

Results are reported as __________________.

---

[1] Source: New York City Division of Special Education's: Test Resource Book.

Technical Information (Procedure for test design; how developed;underlying assumption(s); if and where piloted-number of students and teachers, student ethnicity and SES, etc.; validity including sensitivity to instruction; reliability, how mastery level was determined.)	This test is an outgrowth of ________. The underlying assumption of ________ is __________. The _________ is based on ________. It was piloted on ______ students in ____________________________. __________________ teachers administered the ___________ to students in _____________. Reliability coefficients were reported as ________________. Content validity was determined by _________. p-values ranged from ________. Mastery was determined by the formula_________________. Mastery was determined arbitrarily.
Effective Use/Comments (Adequacy or sufficiency of information to make instructional planning or performance level decisions for individual students; how the test can be used for developing objectives for the IEP (Individualized Educational Program); how it relates with MTEs; comprehensiveness in relation to subject area; specificity of sequencing; evaluative comments; instructional methodology; use with LEP students or linguistically and culturally diverse students.)	The information provided in the ________ is sufficient for making instructional planning decisions or ________________. The information provided in the ________ is sufficient for determining performance levels in the areas of ________. OR Although the _______may be useful for obtaining descriptive information about the student in ________, the test findings are insufficient for making individual instructional planning decision in these areas. It may also be useful for screening purposes. With other measures, the findings may be useful for making instructional planning decisions and .... The information obtained is sufficient for determining performance levels, making instructional planning decisions, and developing ____________. As with other third party scales, _________ is not a direct measure of student performance and consequently judgment of social competence will require direct behavioral information.

The information provided by the ________ is (sufficient) for developing short- and long term objectives on the IEP in the area of________________. This test closely (follows) the sequences of the MTEs. While the manual suggested (reported) ________ minutes, experience indicates_________. While objectives are specific, strategies for obtaining the objectives are not provided. OR In addition to specifying objectives, the manual provides specific strategies regarding instruction. Results with linguistically and culturally diverse students and LEP students should be interpreted cautiously. OR It is not recommended for use with linguistically and culturally diverse students and LEP students. Due to item content, the _________ is not ___________. Its use with ____________ has not been demonstrated. Caution should be used in interpreting results with ______________ due to possible cultural loading of reading passage. Due to the ___________ high verbal loadings, it should be used cautiously with ___________.

References
(use APA style for a complete citation)

Publisher
(Name and address including ZIP code)

Appendix C

SMRT-STEPS SUPPLEMENTAL GUIDELINES FOR REVIEW OF TESTS[1]

Intended Purpose of Test

- What is the description or overview provided by test publisher?
  - The ______ is purported to measure (assess) ______.

Description

- What is the nature of answer sheets and/or test booklets?
  - The ______ consists of reusable (consumable) test booklets.
  - The answer sheets (or booklets) may be hand- or machine-scored.

- Can this test (and subtests) be "customized" or tailored for local use?
  - Supplemental (substitute) subtests, items, and/or objectives are available for ______.
  - No "substitute" subtests, items, or objectives are available.

- How are items clustered?
  - The test has ______ objectives grouped into ______ subtests of ______ items.
  - The test is divided into ______ areas (or subtests) comprised (or consisting) of ______.
  - The test has ______items which are ______.

- How are subtests depicted in test.
  - Similar items are interspersed throughout the test rather than grouped together and appearing in clearly defined subtests.
  - Subtests are clearly defined.

Test Administration

- How are subtests administered?
  - Subtests must be administered in their specified and invariant sequence.
  - It is not specified whether subtests....

- Is the test timed?
  - The ______ is timed and takes ______ minutes to administer.
  - The ______ is untimed but takes approximately ______ minutes to administer.

---

[1] These guidelines were developed by SMRT-STEPS to supplement "Instructions for completing a test review" provided by the New York City Division of Special Education's: Test Resource Guide.

Technical Information

- How was mastery level determined?
    * Mastery (levels were) determined by ______.
    * Mastery (levels were) determined arbitrarily.
    * The manner in which mastery was determined was not specified.

- How does this test relate to curriculum?
    * This test closely follows the sequence of the New York City Board of Education's Minimum Teaching Essentials (MTEs).
    * It is not specified whether this test......

- Was this test developed with teacher input?
    * The test is an outgrowth of ______.
    * ______ teachers administering the ______ to ______ students in ______.
    * The test is based on ______.
    * It is not specified whether the test design was developed with regard to teacher input.

- Was test usefulness reviewed and evaluated by teachers?
    * Teacher judgments were obtained by ______.
    * After administering the test to their students, teachers opinions were elicited.
    * A professional panel consisting of teachers and _____ provided ratings reflecting their opinions of the usefulness of ______.
    * The test design was not developed or re-evaluated as a result of teacher input.

- What steps were implemented to eliminate test bias (i.e., age, cultural, gender, handicapping, racial)?
    * A professional panel reviewed the ______ for potential bias and sensitivity.
    * The test items appear to be free of bias due to ______.
    * The test does not specify whether test items are free of bias.
    * The test does not report whether attempts were made to control cultural, gender,.... bias.

- How was concurrent validity demonstrated?
    * Concurrent validity was demonstrated by comparing test results (performance) of the ______ with similar ______ tests.
    * Concurrent validity, based on comparing the ______ with the ______, indicated a strong relationship with ______ of the ______ subtests.

  - * Concurrent validity ranged from ______ to ______.
  - * Satisfactory concurrent validity was reported.
  - * Correlations between the ______ and the ______ were reported as (high) and were interpreted as being significant (or substantial).

- If reported, how was content validity demonstrated?

- If reported, how was construct validity demonstrated?

- If reported, how was predictive validity demonstrated?

Effective Use/Comments

- Does the test provide norm-referenced interpretations? (Can be in Technical Information)
  - * In addition to providing descriptive information, test results are reported as: age equivalents, grade equivalents, normal curve equivalents percentiles, stanines,.....
  - * In addition to providing (indicating) test scores, the _________ includes descriptive information.

- Can this test be used in program evaluation?
  - * The information provided indicates that this test is appropriate for program (curriculum) evaluation.
  - * ________ is sufficient for determining program effectiveness.

- Can this test be used for developing objectives for the Individual Educational Programs (IEP).
  - * The information provided in the ______ is sufficient for developing short- and longterm objectives on the Individualized Educational Program (IEP) in the area of ______.

- Are specific strategies regarding instruction provided?
  - * In addition to specifying objectives, the manual provides specific strategies regarding instruction.
  - * While objectives are specific, strategies for obtaining the objectives are not provided.
  - * The manual does not provide specific strategies regarding instruction.

- Has it been demonstrated that this test can be used with linguistically and culturally diverse students, Limited English Proficient (LEP) students, handicapped populations?
    * Results with ______ should be interpreted cautiously.
    * It is not recommended for use with ______.
    * Due to item content, ______.
    * Due to test procedure, ______.
    * Its use with ______ has not been demonstrated.